



Assembling, interpreting and contextualizing cellular morphological profiles

Doctoral thesis at the Medical University of Vienna for obtaining
the academic degree

Doctor of Philosophy

Submitted by

Loan Vulliard

M.Sc.

Supervisor:

Univ.-Prof. Dipl.-Phys. Dr. Jörg Menche

Department of Structural and Computational Biology, Max Perutz Labs,
University of Vienna

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences

Vienna, 05/2022

Science was many things, Nadia thought, including a weapon with which to hit other scientists.

Kim Stanley Robinson,
Red Mars.

Declaration

This document presents original research conducted by the author, Loan Vulliard, at the Research Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM) and the Department of Structural and Computational Biology of the University of Vienna, a part of the Max Perutz Labs Vienna, under the supervision of Prof. Jörg Menche. This work led to several publications, including the following first-author articles, included in this thesis:

#1 Book chapter

Vulliard L, Menche J. Complex Networks in Health and Disease. In: Systems Medicine. Elsevier, 2021: 26–33.

The article, in accordance with its license, can be included in the theses and dissertations of its authors for non-commercial uses, allowing its deposition to the university's repository.

#2 Peer-reviewed *Original Paper* published in Bioinformatics

Vulliard L, Hancock J, Kamnev A, et al. BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations. Bioinformatics 2022; 38: 1692–1699. DOI:10.1093/bioinformatics/btab853.

The article is published under an open access license (CC BY-NC 4.0) and can be reproduced for non-commercial purposes.

#3 Peer-reviewed *Resource* article published in Cell Reports

German Y*, **Vulliard L***, Kamnev A, et al. Morphological profiling of human T and NK lymphocytes by high-content cell imaging. Cell Rep 2021; 36: 109318.

* These authors contributed equally.

The article is published under an open access license (CC BY-NC-ND 4.0) and can be reproduced for non-commercial purposes.

All publications can be reprinted as part of this thesis according to the respective policies of their publishers, Elsevier and Oxford University Press.

The contributions of the different authors to these publications are detailed when the respective works and their significance are introduced.

I, the author, declare that unless stated otherwise, all the content of this thesis is previously unpublished and written by myself. A full list of all publications related to my doctoral work is also provided with my curriculum vitae at the end of this thesis.

Table of contents

Declaration	iii
Table of contents	v
List of Figures	vii
Abbreviations	ix
Acknowledgments	xi
Abstract	xiii
Zusammenfassung	xv
1 Introduction	1
1.1 High-throughput screening of cellular perturbations	3
1.1.1 Learning about systems through perturbations	3
1.1.2 Scaling up with biological screens	4
1.1.3 Diversity of perturbations	6
1.2 High-content imaging and profiling	9
1.2.1 Light microscopy	9
1.2.2 Automated image acquisition	11
1.2.3 Successes of HCS	13
1.2.4 Image analysis	15
1.2.5 Morphological profiling	17
1.2.6 Distance between profiles	20
1.3 Network-based modelling of biological systems	23
1.3.1 Studying biological systems as a whole	23

1.3.2	Complex Networks in Health and Disease	26
1.3.3	Network medicine as a framework	34
2	Aims of the thesis	37
3	Results	43
3.1	BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations	43
3.2	Morphological profiling of human T and NK lymphocytes by high-content cell imaging	53
4	Discussion	77
4.1	Future of BioProfiling.jl	78
4.2	Further profiling human lymphocytes	80
4.3	Understanding biological systems through perturbations	81
4.4	Expanding the reach of morphological profiling	84
4.5	Importance of distances to quantify morphological activity	86
4.6	Interfacing morphological profiling and network medicine	89
	Bibliography	93
	Appendix	122
	Supplementary material - BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations	122
	Supplementary material - Morphological profiling of human T and NK lymphocytes by high-content cell imaging	141
	Curriculum vitae	152

List of Figures

1.1	Schematic chart of arrayed and pooled screens.	5
1.2	Typical aims of chemical and genetic screens.	6
1.3	Common steps in high-content imaging experiments.	14
1.4	Distances accounting for dispersion to compare two statistical distributions.	21
2.1	Potential contribution of morphological profiling and network medicine to research in molecular biology.	41
4.1	First part of the follow-up project dealing with genetic perturbations.	82
4.2	Second part of the follow-up project dealing with mixed chemical and genetic per- turbations.	84
4.3	Schematic depiction of how different biological and technical parameters may affect the dispersion of morphological profiles.	88
4.4	Interaction of high-content screening, morphological profiling and network medicine.	89

Abbreviations

- CeMM: research center for molecular medicine of the Austrian academy of sciences
- CRISPR: clustered regularly interspaced short palindromic repeats
- DNA: deoxyribonucleic acid
- HCI: high-content imaging
- HCS: high-content screening
- IS: immunological synapse
- MAD: median absolute deviation
- MHC: major histocompatibility complex
- mRNA: messenger RNA (see *RNA*)
- PCA: principal component analysis
- PPI: protein-protein interaction
- RNA: ribonucleic acid
- sgRNA: single-guide RNA (see *RNA*)
- shRNA: short hairpin RNA (see *RNA*)
- TIFF: tagged image file format
- UMAP: uniform manifold approximation and projection

Acknowledgments

It takes a village to raise a child. PhD titles are awarded to individuals, yet only collectives can create the fertile ground for scientific achievements. I was especially lucky to be surrounded by kind and supportive people, and any merit to the work presented in this thesis must be attributed to them.

First, I want to thank Jörg for his mentorship, his enthusiasm, and for always striving to foster a diverse and collaborative team. In turn, this team was always nice and helpful in many ways. I'm grateful to all the *mobsters*, past and present, for sharing discussions, tips, concerns, meals and code snippets, among others. Thank you Anna for the fun philosophical discussions and the book recommendations. Thank you Celine for always being a great team (and band) player. Thank you Ize for keeping us organized and for the L^AT_EX template. Thank you Joel for sharing this adventure since our first day in the group, and for always being ready to assist, whether with linear algebra or heavy furniture. Thank you Michi for the general pearls of wisdom and for showing us how to get a PhD while still enjoying what you do. Thank you Raphael for bouncing ideas off and tennis balls back at me, and for staying both patient and exigent. Thank you Salvo for the extensive chess training, the pizza and movie reviews, and the piano accompaniment. Thank you Karl, Felix and Daniel for your keen eyes.

I had the pleasure to meet and collaborate with great people outside of the Menche group as well. I believe that science greatly benefits from the combination of various skills and from multidisciplinary approaches, and I want to thank you all for giving me a chance to be a part of such projects. Thanks to Joanna Loizou, Vanja Nagy, Loïc Dupré and their teams, we were able to plan amazing experiments and include some much-needed biological relevance into what we did. Thank you for bearing with me and being passionate about your research, which kept me motivated from the onset of our collaboration and to this day. It was an honor and a pleasure to work with all of you, and especially Yolla, Chris and Joana, who not only kept doing their best and ran awesome experiments, but also educated and entertained me a great deal along the way.

I also want to acknowledge the support I received from everyone first at CeMM and then at the Max Perutz Labs, in particular from the admin teams. Working in these two great institutes, taking care of their scientists and providing them with the best material and intellectual environment possible, felt like winning the lottery twice in a row. Thank you to my thesis committee, Vanja Nagy, Martin Posch and Berend Snijder, for providing much-appreciated guidance. You, together with many of the former mentors, teachers and professors I had the chance to meet, expanded my horizon and convinced me of the importance and relevance of scientific research. I want to give a special thanks to all my friends, who make life worth living and studying, and in particular to Paul, Vincent and the fellow PhD students from Lyon for their strong moral support. Last but not least, I cannot thank my family enough, without whom I would literally not be here today. Merci !

Abstract

With the advent of modern high-throughput techniques, research in molecular biology becomes increasingly quantitative and produces large and complex datasets. Progress in imaging modalities and automation enabled new microscopy-based assays following this paradigm. In particular, high-content imaging screens propose the systematic observation of the response of cells challenged with diverse perturbations. A major open challenge resides in making the most out of these experimental datasets and extracting meaningful and actionable biological knowledge. Here, we explore what can be achieved in image-based screens and how to benefit the most from the spatial and multidimensional information they provide. Our approach of choice starts by compiling multiple descriptors of cellular morphology in integrative profiles used to represent each experimental condition, which is known as morphological profiling. We develop new ways to assemble, curate and work with these profiles. We also observe that combining the profiles with follow-up experiments or integrating them with prior knowledge represented as biological networks is essential to contextualize the results of high-content imaging screens and derive novel insight.

In this cumulative thesis, we first review the state of the art of the literature pertaining to high-content screening analyses and their applications to systems medicine. We introduce the main concepts and technologies employed to test libraries of cellular perturbations, detail how informative morphological profiles can be assembled from cell images, and justify their importance as a way to study biological systems at scale. We also describe how high-content screening results allow these systems to be represented as biological networks, suitable for exploratory and integrative data analysis techniques. Second, we introduce BioProfiling.jl, a computational toolkit that we designed to efficiently work with morphological profiles of cellular measurements. This simplifies the filtering and transformation of such measurements and offers new ways to quantify effects on cell shape and organization. Our main application shows how the tool is able to curate informative profiles describing the mechanism of action of small molecules and provides insight into the complementarity of the morphological changes they induce and the proteins they target. Third, we demonstrate a concrete use case of morphological profiling by studying immune cells. We perform a morphological study of the immunological synapse of cytotoxic lymphocytes. In this study, we profile human T and NK cells and observe the three-dimensional reorganization of actin cytoskeleton and lytic granules. This allowed us to discriminate between several actin defects and also guided further experiments to study dynamic and functional aspects of the immunological synapse formation and degranulation processes that could not be directly assessed via high-content imaging. Lastly, we discuss the outcome of our studies, remaining limitations, and future research directions opened by the recent progress we and others made with respect to morphological profiling. This includes follow-up image-based screens in our group to assemble and study perturbation networks, enabled in part by the work described in this document.

In conclusion, the published studies compiled in this thesis expand what can be achieved with cellular morphological profiles and propose new ways of contextualizing results. Notably, this is achieved by modelling the relations revealed by the profiles with machine learning and network models. Overall, morphological profiling in combination with image-based screens of large perturbation collections can now be used as an informative and scalable approach to shed light on cellular organization. Morphological profiling can thus be leveraged to systematically interrogate and compare the effect of cellular perturbations in a way comparable and compatible with other well-adopted high-throughput experimental modalities, commonly referred to as “OMICS” technologies. As opposed to more targeted assays, one can simultaneously probe a large number of genes, proteins, pathways or mechanisms of action, by characterizing perturbations and how they relate in a single experiment. We progressively transition from a static view of cells and organisms as the simple collection of isolated processes to a more complete picture of properties and phenotypes emerging from complex molecular connectivity patterns. In turn, this updated picture is ideally fitting within the framework of systems medicine, with applications ranging from personalized treatments to improved diagnostics.

Zusammenfassung

Mit dem Aufkommen moderner Hochdurchsatztechniken wird die Forschung in der Molekularbiologie zunehmend quantitativ und erzeugt große und komplexe Datensätze. Die Fortschritte bei den bildgebenden Verfahren und der Automatisierung ermöglichten neue Mikroskopie-basierte Untersuchungen, die diesem Paradigma folgen. Insbesondere bieten bildbasierte High-Content Screens die systematische Betrachtung von Zellreaktionen an, die unterschiedlichen Perturbationen ausgesetzt werden. Eine große Herausforderung besteht darin, wichtigen Informationen aus diesen experimentellen Datensätzen herauszuholen und aussagekräftige sowie umsetzbare biologische Erkenntnisse zu gewinnen. Hier untersuchen wir, was mit bildbasierten Screens erreicht werden kann und wie man von den räumlichen und multidimensionalen Informationen profitieren kann, die sie liefern. Der von uns gewählte Ansatz beginnt mit der Zusammenstellung mehrerer Deskriptoren der Zellmorphologie in integrativen Profilen, die zur Darstellung jeder einzelnen Versuchsbedingung verwendet werden. Dieser Ansatz ist als morphologisches Profiling benannt. Wir entwickeln neue Wege, um diese Profile zusammenzustellen, vorzubereiten und mit ihnen zu arbeiten. Weiter stellen wir fest, dass das Kombinieren der Profile mit Folgeexperimenten oder die Integration von vorherigem Wissen, etwa in Form biologischer Netzwerke, notwendig ist, um die Ergebnisse von High-Content Analysen zu kontextualisieren und neue Erkenntnisse zu gewinnen. In dieser kumulativen Dissertation geben wir als Erstes einen Überblick über den aktuellen Stand der Literatur in Bezug auf High-Content Analysen und ihrer Anwendungen in der Systemmedizin. Wir stellen die wichtigsten Konzepte und Technologien vor, die zum Testen von zellulären Perturbationen verwendet werden, und erläutern, wie informative morphologische Profile aus Zellbildern erstellt werden können. Die Bedeutung dieser Profile wird als Mittel zur Untersuchung biologischer Systeme in einem kontextbasiertem Maßstab begründet. Weiters erläutern wir, wie es solche experimentellen Ergebnisse ermöglichen, diese Systeme als biologische Netzwerke darzustellen, die für explorative und integrative Datenanalysetechniken geeignet sind. Zweitens stellen wir BioProfiling.jl vor. Diese Softwarelösung haben wir für die effiziente Arbeit mit morphologischen Profilen von zellulären Messungen entwickelt. Dies vereinfacht die Filterung und Umwandlung solcher Messungen und bietet neue Möglichkeiten zur Quantifizierung der Auswirkungen auf die Zellform und -organisation. Unsere Hauptanwendung zeigt, wie das Programm informative Profile erstellt, die den Wirkmechanismus kleiner Moleküle beschreiben. Diese Profile geben Einsicht in die Zielproteine der chemischen Verbindungen und in die Komplementarität der von ihnen ausgelösten morphologischen Veränderungen. Drittens zeigen wir durch die Untersuchung von Immunzellen einen konkreten Anwendungsfall der morphologischen Profilerstellung. Wir führen eine morphologische Studie der immunologischen Synapse von zytotoxischen Lymphozyten durch. In dieser Studie erstellen wir Profile menschlicher T- und NK-Zellen und beobachten die dreidimensionale Reorganisation des Aktinzytoskeletts und der lytischen Granula. Auf diese Weise konnten wir zwischen verschiedenen Aktin-Defekten unterscheiden und weitere Experimente durchführen, um dynamische und funktionelle Aspekte

der immunologischen Synapsenbildung und des Degranulationsprozesses zu untersuchen, die mit High-Content Bildgebung nicht direkt erfasst werden konnten. Abschließend erörtern wir die Ergebnisse unserer Studien, verbleibende Einschränkungen und zukünftige Forschungsrichtungen anhand jüngster Erkenntnis, die wir und andere in Bezug auf die morphologische Profilerstellung erzielt haben. Dazu gehören weitere bildbasierte Screens zum Aufbau und zur Untersuchung von Perturbationsnetzwerken in unserer Arbeitsgruppe, die zum Teil durch die in diesem Dokument beschriebenen Arbeiten ermöglicht wurden.

Zusammenfassend erweitern die hier veröffentlichten Studien die Möglichkeiten der zellulärer morphologischer Profile und schlagen neue Wege zur Kontextualisierung der Ergebnisse vor. Dies wird insbesondere durch die Modellierung der durch die Profilerstellung enthüllten Beziehungen mit maschinellem Lernen und Netzwerkmodellen erreicht. Insgesamt kann die morphologische Profilerstellung in Kombination mit bildbasierten Screens großer Sammlungen von Perturbationen fortan als informativer und skalierbarer Ansatz genutzt werden, um die zelluläre Organisation zu analysieren. Die morphologische Profilerstellung kann somit genutzt werden, um die Auswirkungen zellulärer Perturbationen systematisch auf eine Art und Weise zu untersuchen und zu vergleichen, die vergleichbar und kompatibel mit anderen experimentellen Hochdurchsatzverfahren ist, die gemeinhin als „OMICS“-Technologien bezeichnet werden. Im Gegensatz zu gezielteren Assays kann man eine große Anzahl von Genen, Proteinen, Stoffwechselwegen oder Wirkmechanismen gleichzeitig untersuchen, indem man die Perturbationen und ihre Beziehungen in einem einzigen Experiment charakterisiert. Wir wechseln allmählich von einer statischen Betrachtung von Zellen und Organismen als einfache Ansammlung isolierter Prozesse zu einem vollständigeren Bild, in dem Eigenschaften und Phänotypen aus komplexen molekularen Vernetzungsmustern sich ergeben. Dieses aktualisierte Bild passt wiederum ideal in den Rahmen der Systemmedizin, deren Anwendungen von personalisierten Behandlungen bis hin zu verbesserter Diagnostik reichen.

1

Introduction

Over the years, the growing molecular understanding of cell biology helped to identify several pillars supporting health. These pillars can overall be grouped in three main domains: (i) maintenance of homeostasis, (ii) response to stress and (iii) spatial compartmentalization (López-Otín & Kroemer, 2021). To maintain homeostasis, biological systems require a careful balance between diverse cellular components and their interactions. Thus, these systems need to be studied in their entirety instead of looking at their individual components in isolation, which is the main promise of systems biology (Ideker et al., 2001). In particular, molecules and their interaction can conveniently be represented as biological networks (Barabási & Oltvai, 2004). The response to stress covers the ability of all organisms to adapt to changes and to recover from disturbance (Lee et al., 2014). Exposures, diseases, or even therapeutic interventions, constitute perturbations of the delicate molecular equilibrium underlying homeostasis, either supporting or disrupting it. Understanding these perturbations and accordingly how to navigate the landscape of cell states to go from disease to health is at the core of the field of network medicine (Loscalzo et al., 2017). The concept of spatial compartmentalization is also critical, as direct interaction is only possible if molecules are close enough to bind and more generally, physical proximity is a marker of shared function and of the modular organization of the cell (Qin et al., 2021).

Few experimental setups allow to both preserve spatial information and provide a systematic way to assess the effect of large collections of perturbations of biological processes. To this end, an approach named high-content imaging screening allows for probing the morphology of a model system in multiple experimental conditions (Boutros et al., 2015). Typically, cells are seeded in microplates and an automated fluorescence microscope provides thousands of images revealing the spatial distribution of proteins and organelles. Coupled with image analysis techniques, it is possible to extract measurements describing the morphological changes induced by perturbations, such as drugs or genetic knockouts. These phenotypic measurements can be combined into morphological profiles that can be compared and integrated with additional

information about the conditions they describe (Bougen-Zhukov et al., 2017; Caicedo et al., 2017). Successful uses of this approach result in quantitative representations of the studied experimental conditions that reveal novel information about cellular perturbations. In this way, morphological studies contributed to the identification and characterization of chemical modes of action (Ziegler et al., 2021), gene functions (Moffat et al., 2006) and perturbation interaction profiles (Breinig et al., 2015; Caldera et al., 2019; Heigwer et al., 2018).

The success of such analyses is conditioned on the compilation of informative morphological profiles, which presents both technical and scientific challenges (Bougen-Zhukov et al., 2017). For instance, how to best choose morphological descriptors, integrate them despite their heterogeneity, account for experimental confounders, and which statistical tools are appropriate for their analysis? Accordingly, we dedicate this document to new ways to assemble, interpret and contextualize morphological profiles. This constitutes the central aim of this thesis, which is further detailed and motivated in chapter 2. To achieve this goal, we hereby compile articles, all devised in the course of the doctoral work associated with this thesis. These articles describe improvements and applications of cellular morphological profiling, and participate in establishing that the combination of imaging screens, morphological profiles and network medicine covers all three categories of hallmarks of health, with the potential to provide novel biological insight.

The doctoral work described in this document was performed in the team led by Prof. Jörg Menche, which specializes in the development of network-based analysis methods, and their applications to biomedical research. In particular, a long-term goal pursued by our research group is to derive an arithmetic of cellular perturbations, which would be capable of describing the response of cells to multiple internal and external stresses and exposures. Our group previously analysed a high-content screen of drug combinations and linked different modes of interaction to properties of drug targets on the protein-protein interaction network (Caldera et al., 2019). In the future, larger and more comprehensive screens and could allow us to further uncover the basic principles underlying cellular response to perturbations. Enabling ambitious expansions of this work, such as the follow-up projects described in section 4.3, was an additional motivation to make morphological profiling simpler and more powerful.

This first chapter defines the key concepts at the core of the compiled articles and gives background information to understand and contextualize them. As such, the material provided here prioritizes information that helps in motivating and explaining what we achieved. A first section is dedicated to molecular perturbations as a means to study cell responses, and more specifically in the context of screening experiments. A second section covers more specifically microscopy-based assays, and presents the typical outline of a high-content imaging screen, including the main computational analysis steps. Finally, a third section discusses the topics of systems and network biology, which are essential for the contextualization of system-wide morphological analyses, and their translational applications via network medicine.

1.1 High-throughput screening of cellular perturbations

In this section, we briefly present how perturbing biological systems may shed light on the molecular mechanisms underlying these systems. Experimentally, multiple perturbations can be used individually or in combination to systematically study cellular response in large-scale screens. We discuss the two main designs of biological screens, namely pooled and arrayed screens, along with their strengths and characteristics. Furthermore, we mention the different types of perturbations typically used in screens, and how they can answer different biological questions.

1.1.1 Learning about systems through perturbations

Biological research historically followed a reductionist approach, generating data and insight on increasingly smaller and precise phenomena with the hope that it would in turn help to elucidate phenotypic effects observed at a larger scale (Lazebnik, 2002). Yet, a better resolution into the components of a biological system also implies a larger number of components to describe. This means that biologists constantly have to face the challenge of scaling up their experimental setups. For instance, for a disease or phenotype of interest, one could wonder how different organs or tissues are involved, or going more in-depth what the role of every protein or residue is. Each change of biological resolution is associated with a change in the magnitude of the number of its components, going from around 8×10^1 types of organs to 3×10^{13} cells in a human body, each with 2×10^4 types of proteins encoded by a 6×10^9 -nucleotide-long genome (Nurk et al., 2021; Sender et al., 2016). On top of this complexity, the combinatorial explosion comes into play, with potential pairwise and higher-order interactions between each of these elements (Kuzmin et al., 2018). Observations made in a given environment or specific genetic background may not hold in another setting. Thus, it is typically necessary to study biological properties in multiple conditions. For instance, it is well established that the sensitivity to chemical inhibition of molecular targets is largely influenced by genetic factors, at the level of both cells and patients (Ahmed et al., 2016; Piotrowski et al., 2017).

Being able to study the entirety of such systems requires a change in analytical approaches, particularly visible with modern molecular biology methods such as genome-wide association studies or next-generation sequencing. The development of these techniques is indeed tightly linked to the expansion of statistical frameworks dedicated to the interpretation of such experimental results (Law et al., 2014; Ritchie et al., 2015; Uffelmann et al., 2021). It also requires an optimization of experimental setups to achieve a reduction in cost and an increase in automation and throughput. The process of testing as many perturbations as possible in a systematic manner on a given biological model is referred to as high-throughput screening, and was pioneered decades ago by the pharmaceutical industry, initially to accelerate drug

discovery and development (Pereira & Williams, 2007). Typically, multiple candidate drugs would be tested to identify “hits” with a desired property in a simple model system. These hits would be further validated and optimized before being considered “lead compounds” and undergoing clinical trials (Malo et al., 2006). One would also differentiate between primary screens, intended to narrow down a large number of candidates to a smaller number of hits, and secondary screens which would be more extensive and aim to validate the robustness of the hits identified in the primary screens. Other screening approaches, introduced in the following subsection, favor lower-throughput readouts as they may be more informative or more relevant for a given application. Screens can also be performed on a wide range of biological systems, including in three-dimensional culture and whole organisms for species such as *Danio rerio* (zebrafish) and *Caenorhabditis elegans* (roundworm), which provides valuable information when studying development or growth phenotypes (Han et al., 2020; Lukonin et al., 2020; Mondal et al., 2016; Wählby et al., 2012; Westhoff et al., 2020). In the rest of this thesis, we focus on the more common application to two-dimensional cell culture, including both immortalized cell lines and material directly derived from donors and patients.

1.1.2 Scaling up with biological screens

Overall, biological screens aim to achieve an optimal balance between cost, scale, robustness and information content. They are usually split in two main categories displayed on Figure 1.1: arrayed and pooled screens.

In arrayed screens, each perturbation is tested in a separate well of a microplate. Each well accommodates for a perturbation or for a negative control, providing a baseline of how unperturbed cells behave, which typically is a solvent in a drug screen, or a non-targeting control in a genetic screen. In these experiments, plates most often have 96 or 384 such wells, which are used as independent “test tubes”. This allows the physical segregation of the growth medium, and several compounds or concentrations can be used without interfering with each other. This comes at the expense of practicality, and requires dedicated resources, such as multichannel pipettes or automated liquid dispensers. Experiments performed in wells with a limited volume are also prone to plate-layout biases, due to increased medium evaporation at the edges, although the effect can be alleviated by the use of a randomized layout (Caicedo et al., 2017; Malo et al., 2006). However, when these challenges can be overcome and after accounting for technical confounders, arrayed screens offer a powerful and straightforward way to systematically test the effect of various perturbations.

In pooled screens, one makes use of the knowledge of the initial and final composition in a single well or dish where multiple perturbations are pooled, and infers the impact of the perturbations through the changes that occurred (Doench, 2017). This is often cost-effective and has multiple advantages. It enables the profiling of multiple perturbations in a single dish and involves less repetitive tasks than arrayed screens, in which each step needs to be replicated once for

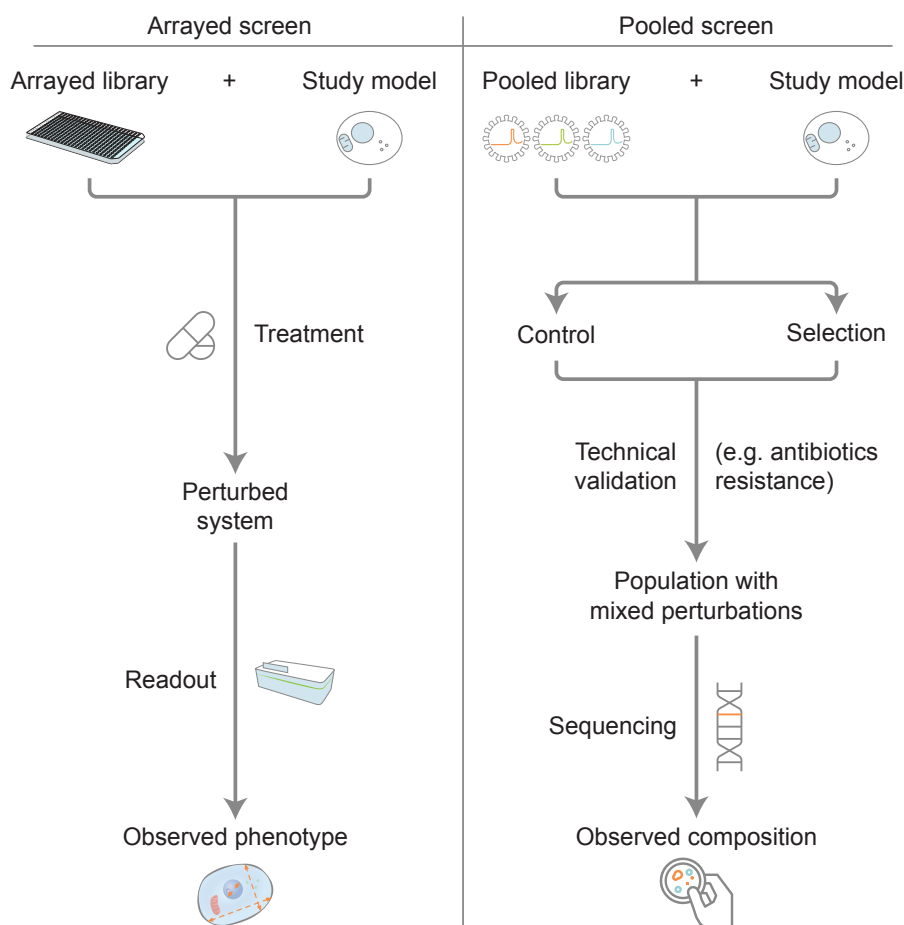


Figure 1.1: Schematic chart of arrayed and pooled screens. Arrayed screens typically involve seeding cells in the wells of a microplate, each representing an experimental condition. A readout is later obtained in each well, allowing a side-by-side comparison. Pooled screens usually expose a cell population to a pooled perturbation library. The relative composition of the final cell population is most often assessed by high-throughput sequencing, and informs on the impact of the perturbations compared to a reference treatment.

each microwell in use. Despite these strengths, pooled screens also have weaknesses. A main challenge in setting up a pooled screen lays in preparing a pooled library of perturbations, in which the initial composition is precisely known. A selection step follows, corresponding to the phenotype of interest, such as resistance to a compound or expression of a chosen biomarker. In genetic screens, this selection can also serve as a control to ensure that the remaining cells were successfully transfected, for instance via the expression of a puromycin resistance gene delivered together with the material enabling the genetic editing. As opposed to arrayed screens where the perturbation used in each well is known, pooled screens require a way to tell which modification was received by each cell in the final population. For this, the experimental readout needs to be able to discriminate between perturbations. For instance, short RNA sequences delivered with a viral construct can serve as barcodes to identify each perturbation, or in the case of CRISPR screens, the sequence of the guide RNA can be amplified directly. After targeted sequencing, it is therefore possible to quantify if the relative abundance of a given perturbation

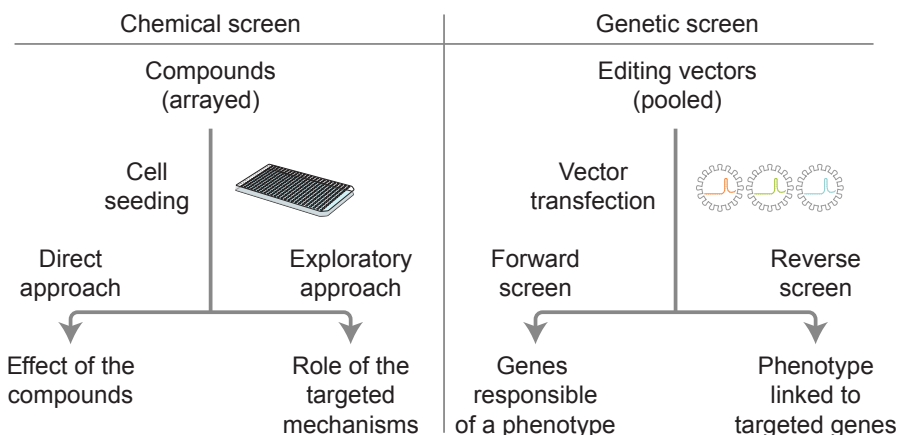


Figure 1.2: Typical aims of chemical and genetic screens. In a typical chemical screen, cells are exposed to drugs or bioactive molecules, either to study cellular response or to learn about the process disrupted by the chemicals. The two main purposes of genetic screens are the identification of genes underlying a phenotype of interest (forward screen), and the study of the phenotypic consequences of a gene disruption (reverse screen).

increased, in case of a positive selection, or decreased, due to negative selection. Of note, the latter is harder to detect and requires a larger sample or effect size. For instance, ineffective or monoallelic edits might cause the incomplete penetrance of a deleterious genetic edit. This means that the unaffected cells could still be present after selection, as the resistance gene would be expressed, but the amplitude of the decrease in abundance of the target gene could be small. On the contrary, a beneficial edit could anyway lead to a clonal advantage and a large overrepresentation in the final genetic pool, and be associated with a large effect size (Sharma & Petsalaki, 2018). Furthermore, a large population is needed to avoid bottleneck effects, for which a subpopulation with low abundance would completely disappear only due to random fluctuations (Doench, 2017). Traditionally limited to low-content and one-dimensional readouts such as differential abundance, creative approaches have also been proposed in recent years to bring pooled screens closer to the flexibility offered by arrayed screens (Datlinger et al., 2017; Dixit et al., 2016; Feldman et al., 2019; Wang et al., 2019; Wheeler et al., 2020).

All screens presented in this thesis follow the arrayed approach. We based our choice on the convenience and directness of the method, and on the access to the chemical screening facility at CeMM from which we and our collaborators benefited. The two research papers compiled in the result section thus deal with arrayed high-content imaging (HCI) screens. As of 2022 and despite gradual decrease in the cost of sequencing, imaging indeed provides a more cost efficient readout for arrayed screens and comes with additional advantages listed in subsection 1.2.

1.1.3 Diversity of perturbations

There are multiple types of perturbations affecting cellular phenotypes that can be explored via screening, allowing for flexible use in both basic and applied research (Bougen-Zhukov et al.,

2017). Generally, perturbations fall within two main classes, (i) extrinsic and (ii) intrinsic, as they can either affect cells from the outside and trigger a response, or directly alter the inner components of the cell. Most often, this takes the form of chemical and genetic perturbations, respectively. We will now explore how both classes of screens are designed and what they have to offer, starting with chemical screens and continuing with genetic screens.

Chemical perturbations correspond to the use of compounds as a way to modulate the state, activity and function of the cell. They range from natural products to collections of small molecules synthesized to maximize structural diversity (Barnes et al., 2016; Bray et al., 2017; Bryce et al., 2019). These perturbations chiefly serve two distinct purposes. They might directly be of interest to the investigators, for instance when a drug candidate is studied to be used as a treatment, or be good models to learn about the processes underlying responses to such perturbations. For example, a compound with a known mechanism of action that inhibits an enzyme could be used to learn about the role of the corresponding pathway. This dichotomy between a direct and an exploratory approach is illustrated in Figure 1.2. Both use cases are covered in the publications compiled in this thesis. The first article, reproduced in section 3.1, is directly describing the role of the selected perturbations. It includes a chemical screen where many compounds are approved or under investigation for disease treatment. The analysis compares the effect of compounds on cellular morphology with an emphasis on the relation between their established mechanism of action and their propensity to induce morphological changes. In contrast, in the second article, presented in section 3.2, the chemical perturbations serve to explore biological conditions of interest. There, different compounds are used to probe the association of morphological changes with the inhibition of different steps of actin cytoskeleton organization, including F-actin polymerization and branching.

The second main class of screens relies on genetic perturbations, which provide a key tool for functional genomics. These screens can be further subdivided into forward and reverse screens (Figure 1.2). Forward screens aim to identify the genotype underlying a phenotype of interest, for instance the genetic causes of a functional defect. Reverse genetic screens start from a genetic change to identify its phenotypic impact. Put simply, one can annotate the role of genes by knocking them out one at a time. Genetic perturbations encompass different types of alterations of either DNA or RNA. First, the genome can be directly edited to induce mutations, insertions or deletions. Molecular biologists have long used type II restriction enzymes isolated from archaea and bacteria to cleave DNA at positions of interest. This massive adoption by the research community led Werner Arber, Daniel Nathans and Hamilton Smith to be rewarded with the Nobel Prize in Physiology or Medicine in 1978 (Pingoud et al., 2014). More recently, the artificial expression of CRISPR systems in combination with the endonuclease enzyme Cas9 has emerged as a tool of choice for genetic screens. In turn, this was also recognized with a Nobel prize in Chemistry awarded to Emmanuelle Charpentier and Jennifer Doudna in 2020. Naturally present as an immune defense mechanism against phage infections in 50% of bacteria and 90% of archaea, CRISPR-Cas systems can also be synthetically introduced in mammalian cells (Wright et al., 2016). This method induces programmable

mutation-prone double strand breaks in genomes, leading to near-complete gene knockouts via nonsense-mediated mRNA decay (Jinek et al., 2012; Wang et al., 2014). This genetic engineering approach only requires the design of a short custom RNA sequence, called single-guide RNA (sgRNA), to target a genetic location of interest, and displays a significant improvement in editing rate compared to previous technologies such as short hairpin RNA (shRNA) (Doench et al., 2016; Shalem et al., 2014). While the understanding of the DNA damage and DNA repair mechanisms involved in CRISPR-mediated gene knockouts deepens, it is also becoming clearer that efficient knockouts are not possible in all biological contexts. Among other limitations, they are tissue-specific when used in vivo (Ferreira da Silva et al., 2021). Therefore, technological improvements will be needed to expand the potential applications of CRISPR screens. Other approaches do not rely on mutagenesis but rather aim to silence or enhance the expression of a gene. Gene knockdown experiments, in which expression levels are artificially reduced, are predominantly performed at the level of nucleic acids. DNA transcription can be repressed, which is the case in CRISPR interference experiments, where an inactivated mutant of the Cas9 protein is used in combination with sgRNA to selectively block transcription initiation or elongation (Qi et al., 2013). Conversely, gene overexpression can be induced by fusing the inactive Cas9 with transcriptional activators in CRISPR activation experiments (Gilbert et al., 2014). RNA transcripts can also be targeted in RNA interference experiments, for instance through the delivery of custom vectors which trigger a cascade leading to the cleavage of a targeted transcript (Brummelkamp et al., 2002). These perturbations are typically induced thanks to a vector transfected to the cells, either in a stable or in a transient manner. This is often achieved chemically via electroporation, or using lentiviral transduction (Bock et al., 2022). All these technologies have already been leveraged in large-scale screens, sometimes covering the entire genome (Boutros et al., 2006; Kampmann, 2018). Finally, perturbations acting at the protein-level are typically induced with chemical inhibitors, and thus typically follow the experimental outline of compound screens described above. Targeted protein degradation could lead to a more systematic way to selectively remove proteins, which could then be used in screens. However, this has not been achieved yet, and works in this field focused so far on genetic screens to study and optimize a limited number of degraders at a time (Scholes et al., 2021).

Finally, screening libraries are not limited to either genetic or chemical perturbations. Some studies combine both simultaneously (Breinig et al., 2015; Gapp et al., 2016; Piotrowski et al., 2017), while others explore different types of perturbations altogether. For instance Bruch, Giles and colleagues studied the influence of multiple microenvironments on the effectiveness of drugs used to treat chronic lymphocytic leukaemia (Bruch et al., 2021). In subsection 1.2.3, additional examples of screens with purposes ranging from basic to translational research are also introduced.

1.2 High-content imaging and profiling

Cell morphology has been found to be directly indicative of specific diseases. For instance, a disruption of polarity can be observed in epithelial cancer cells (Muthuswamy & Xue, 2012), and interactions between pathogens and the actin cytoskeleton of host cells often lead to visible changes (Haglund & Welch, 2011). Moreover many genetic defects are associated with phenotypic changes observable via quantitative imaging experiments, notably inborn errors of immunity (Brigida et al., 2018; Dupré et al., 2002; Pfajfer et al., 2018) and neurological disorders (Fell et al., 2021; Sharma et al., 2013; Zamboni et al., 2018). Accordingly, imaging technologies are especially suitable to study cell biology in health and disease, and are at the core of the experimental effort described in the studies compiled in this thesis. In this section, we introduce imaging modalities and their applications to large-scale screens. We start with a presentation of light microscopy and its main promises. We then cover how automated microscopy platforms enabled a new screening modality, high-content screening, providing multiple morphological descriptors as readout. We compile some successful applications of this approach which highlight the diversity of use cases in which high-content imaging is employed. We then proceed with an overview of the computational analysis of the acquired images, and how morphological measurements can be compiled in profiles describing different phenotypic cell states across experimental conditions. The final part of this section deals more specifically with the ways in which differences between morphological profiles can be quantified. This is a prerequisite to define which perturbations impact cellular morphology significantly, and is essential to the understanding of the work presented later in section 3.1.

1.2.1 Light microscopy

Microscopy has been a canonical way to characterize living organisms since the first prototypes of microscopes built by Antonie van Leeuwenhoek and Robert Hooke, which led to the first observations of microorganisms in the late seventeenth century. Since then, technical improvements went hand in hand with scientific advancement. In its most simple form, a microscope is a tool allowing to visualize an object at a larger size while preserving most details. Typically, the magnification occurs through a system of optical lenses. A source of light illuminates the study object. This object, the sample, absorbs and deflects the light based on its structure and composition. The objective lens of the microscope then gathers and focuses the light effectively reaching it, allowing the observation of the sample at a given magnification factor. This basic principle is still at the core of modern-day light microscopy, which offers a sufficient resolution to observe subcellular structures without strongly perturbing the imaged cells (Thorn, 2016). Improvement in manufacturing techniques led to a better quality and a tighter setup of all microscope parts, while progress in optics led to enhanced microscope designs. Together, the

three key aspects of microscopy, namely magnification, resolution and contrast, were improved. A major development, with a large impact on the biological applications of microscopy was the adoption of fluorescence microscopy. This scientific revolution was kick-started by the identification of a green fluorescent protein (GFP), naturally present in the jellyfish *Aequorea victoria* (Shimomura et al., 1962). Years later, GFP became an essential part of the molecular biology toolkit, for instance serving as a marker for gene expression (Chalfie et al., 1994). For their work on GFP, Osamu Shimomura, Martin Chalfie and Roger Tsien were awarded a Nobel Prize in Chemistry in 2008. In brief, fluorophores, also called fluorescent dyes or stains, are used to tag molecules of interest. Typically, this is either done by genome engineering to fuse the dye with a protein of interest, or by immunofluorescence, using fluorescently-labeled antibodies. The fluorophores are then excited with a laser set at a given wavelength and consequently emit light at another fixed emission wavelength, which will in turn be captured by the microscope objective. While this may appear at first as a constraint compared to the label-free approach, this actually opened a whole new range of avenues for molecular biology, by making it possible to selectively highlight structures or proteins of interest (Lichtman & Conchello, 2005). Light only needs to be detected at a specific wavelength, given by physical properties of the fluorophore, which has the additional advantage of being less noisy than bright-field microscopy. If they have distinct characteristic wavelengths, several dyes can also be used to stain several classes of molecules at the same time.

Several well-adopted protocols make use of fluorescence microscopy as part of more complex experimental designs for multiple applications. This includes single-molecule fluorescent in situ hybridization (FISH) and multiplexed error-robust FISH to segment individual mRNA transcripts (Chen et al., 2015; Lyubimova et al., 2013), or the use of Förster resonance energy transfer to assess physical proximity between pairs of molecules in the cell (Algar et al., 2019). For both papers, presented in sections 3.1 and 3.2, images were generated using automated fluorescence microscopes. More precisely, in section 3.2, we used a spinning disk confocal microscope, which limits the presence of out-of-focus staining intensities, occurring as the light is initially emitted from its source in a cone. Put simply, these microscopes include pinholes that block all rays which would not converge at the imaged focal plane. This technology makes it possible to precisely image samples in three dimensions (Thorn, 2016), which allowed us to study the positioning of lytic granules. One central aspect leveraged in both studies resides in the richness of imaging as a high-dimensional readout. Compared to targeted assays providing a single value per measurement, such as cellular viability or yeast colony proliferation, hundreds of morphological descriptors can be derived from each image, as we detail in subsection 1.2.4, thus permitting more applications and nuances. For instance, it becomes possible to characterize precisely how molecular perturbations combine and influence each other (Caldera et al., 2019; Fischer et al., 2015) or to describe several properties of the cells which would typically be described through multiple assays (Simm et al., 2018; Way et al., 2021a). We explore in further detail how one can benefit the most from the high dimensionality offered by imaging as a readout in sections 4.3 and 4.4.

Light microscopy also has limitations. Ernst Abbe already derived in 1873 the optimal resolution that can theoretically be achieved by a microscope due to diffraction, based on the light wavelength and on a physical property of the microscope called numerical aperture. Only specific conditions and microscopes currently allow experimenters to observe molecules smaller than 200 nm thanks to so-called super-resolution light microscopy (Ji et al., 2008). Other modalities, departing from the traditional light microscopy approach, coexist and offer complementary views of spatial organization of cells at different scales. In particular, electron microscopy proposes to characterize nanoscale cellular structures. This approach usually relies on accelerated electron beams instead of photons to illuminate the sample (Zewail, 2010), and can be used in complement to wide field or confocal microscopes, to combine the global cellular context and details at the single-molecule resolution (de Boer et al., 2015). Using such microscopes, researchers have been able to measure crystal spacing less than 50 pm (Erni et al., 2009). Demonstrating the impact of these technologies, Ernst Ruska was awarded the Nobel Prize in Physics in 1986 for pioneering electron microscopes, and Jacques Dubochet, Joachim Frank and Richard Henderson received the Nobel Prize for Chemistry in 2017 for their work on cryo-electron microscopy. The latter consists in the near-instantaneous freezing of biological samples to protect them and vitrify the surrounding water, and the subsequent use of an electron microscope to resolve molecular structures (Cheng et al., 2017). This method only requires a limited amount of the macromolecule of interest and typically resolves its structure in its native and functional conformation. In this thesis, we focus on static images but other technologies also allow for timelapse imaging of live cells, which is essential to study dynamical processes such as actin-driven motility (Kamnev et al., 2021). Other recent approaches do not rely on either visible light nor electron transmission and could prove complementary, such as DNA microscopy, in which a chemical tagging of transcripts is combined with a proximity-based reconstruction of the cellular arrangement (Weinstein et al., 2019), and expansion microscopy, in which the sample itself is physically made larger (Wassie et al., 2019).

1.2.2 Automated image acquisition

By coupling the microscope objective to a digital camera detector, it is possible to directly record the observed signal. As long as the camera has a pixel size at most half the size of the resolution of the optical image, no information is lost (Bernas, 2005). At this step, the optical image is converted into an array of numbers representing the intensity of light measured at each pixel. This digital image can then be analysed in real time, or stored and analysed at a later time, for instance once the whole imaging component of the experiment has been conducted. In a logical continuation of the high-throughput screening approach introduced in subsection 1.1.1, the automation of cellular imaging fueled its use for assessing the effect of perturbations, and in particular of candidate drugs, in a process called high-content screening (HCS). HCS assays using microscopes to acquire phenotypic measurements are also referred to as high-content

imaging (HCI) experiments. The first commercial systems branded as HCS platforms date back to 1997. In a landmark paper, Giuliano and colleagues rightly predicted that “*new reagents, algorithms, screens, and bioinformatics will increase the power of this approach*” (Giuliano et al., 1997). Of note, the authors of this article were active in both public and private research institutions. From its inception, HCS indeed showed promises for both applied and basic research, and the technology was pioneered by pharmaceutical companies to facilitate drug development. To this day, actors from academia and industry perform HCS experiments and often collaborate to develop the field further (Caicedo et al., 2017; Mullard, 2019). The technological requirements are still expensive, but HCI experiments are progressively getting more affordable and widespread. Several commercial providers, such as PerkinElmer, GE Life Sciences and Thermo Fischer, now offer confocal and wide field microscopes suitable for HCI. On top of short image acquisition time, these instruments typically provide magnification factors up to 100, have three to seven fluorescence channels and allow controlling experimental parameters such as CO₂ levels, temperature and humidity. It is now possible to screen and characterize the effect of tens or hundreds of thousands of compounds in a single experiment (Bray et al., 2017; Bryce et al., 2019) and to stain multiple organelles and cellular structures simultaneously for instance with the Cell Painting assay (Bray et al., 2016). More examples of successful applications of HCS are presented in subsection 1.2.3.

The steps common to most HCI experiments are shown in Figure 1.3, together with the main analytical steps. As described above, many different arrayed cellular perturbations can be used as the starting point of an HCI screen. No matter the chosen perturbations, the process stays relatively unchanged (Boutros et al., 2015). Once the cells are treated or genetically edited, the cells are washed, and stained using fluorescent dyes. As most imaging screens are only evaluating cellular morphology at a single time point, it is often convenient to fix the cells using formaldehyde, as it renders the results independent of the time elapsed between staining and image acquisition, and as it increases the choice of usable dyes. The microscope then proceeds plate by plate, well by well and usually images several fields of view in each well sequentially, without the need for human supervision. Among others, the Opera Phenix microscope from PerkinElmer, used in our study described in section 3.2, offers different strategies to select fields of view. It allows the user to choose a fixed layout to preferentially image the center or the edges of the well, which can be convenient based on how the imaged cells preferentially grow. Alternatively, it can automatically detect and photograph cell clusters based on low-magnification images, which maximizes the number of captured cells but may induce biases, as the sampling process is no longer random and comparable in all wells. This microscope also integrates several cameras to capture the signal in the different fluorescence channels simultaneously to decrease acquisition time. The resulting images, often following the tagged image file format (TIFF) standard, include the measured intensities and basic metadata characterizing the acquisition. These files can then be exported and analysed to characterize cellular morphology or other phenotypes detectable thanks to the fluorescent probes.

Some cell lines are particularly suitable for high-content imaging screening due to their ability

to display specific morphological states matching the mechanism of action of treatments (Kang et al., 2015) or more generally due their expressive morphology, such as U-2-OS which is used in multiple large screens (Bray et al., 2017; Gustafsdottir et al., 2013; Strezoska et al., 2017). This cell line and the previously-mentioned Cell Painting assay were selected to generate a collection of such screens by a large consortium, recently initiated by Dr. Anne Carpenter and collaborators at the Broad Institute of MIT and Harvard and across 10 pharmaceutical companies (Mullard, 2019). The same cell line was used in our study presented in section 3.1 for the advantages it offers and the ability to compare our results to a broad body of literature. U-2-OS cells were kindly provided and expanded by the lab of Ass.-Prof. Joanna Loizou. The imaging was done at the chemical screening facility at CeMM, directed by Anna Koren, and part of the molecular discovery platform headed by Dr. Stefan Kubicek, on the PerkinElmer Operetta HCI system. The facility is conveniently equipped with all the necessary tools to perform and automate HCS experiments, including an acoustic liquid dispenser, a washer, a robot arm to handle the plates, and an incubator to store them until they are imaged.

1.2.3 Successes of HCS

In recent years, HCS contributed to discoveries in cell biology across many studies, especially via morphological profiling, an approach detailed later in subsection 1.2.5. It occupies a unique spot among technologies used for research in molecular biology. Indeed, the number of imaged cells offers an improved statistical power compared to traditional low-throughput light microscopy. Morphological measurements also provide a multidimensional readout that is much more comprehensive than most other high-throughput assays. As HCS was initially intended to facilitate drug discovery (Giuliano et al., 1997), it naturally found many use cases at different stages of drug development and led to the characterization of small molecules and their mechanisms of action in multiple cellular models (Bray et al., 2017; Bryce et al., 2019; Wawer et al., 2014). Applications are, however, reaching far beyond pharmacological research. HCI results allowed researchers to assemble information-rich profiles able to replace multiple costly assays (Simm et al., 2018; Way et al., 2021a) or optimize multiple biological parameters and chemical properties at once (Lin et al., 2021). HCI is especially adequate for describing complex cellular phenotypes. Together with the demonstrated importance of cell shape in pathologies (Barker et al., 2022), HCI appears as a method of choice to study how function and morphology are linked in disease contexts. The flexibility of the experimental setups even makes HCI suitable for translational applications including ex-vivo drug screening on primary cancer patient material (Snijder et al., 2017) and on patient-derived organoids (Betge et al., 2019). An image-based treatment prioritization in hematological cancers has also led to good outcomes in a prospective clinical trial (Kornauth et al., 2021).

Moreover, HCS can be used to address fundamental research questions. Some notable achievements involved the characterization of interactions between multiple chemical or genetic per-

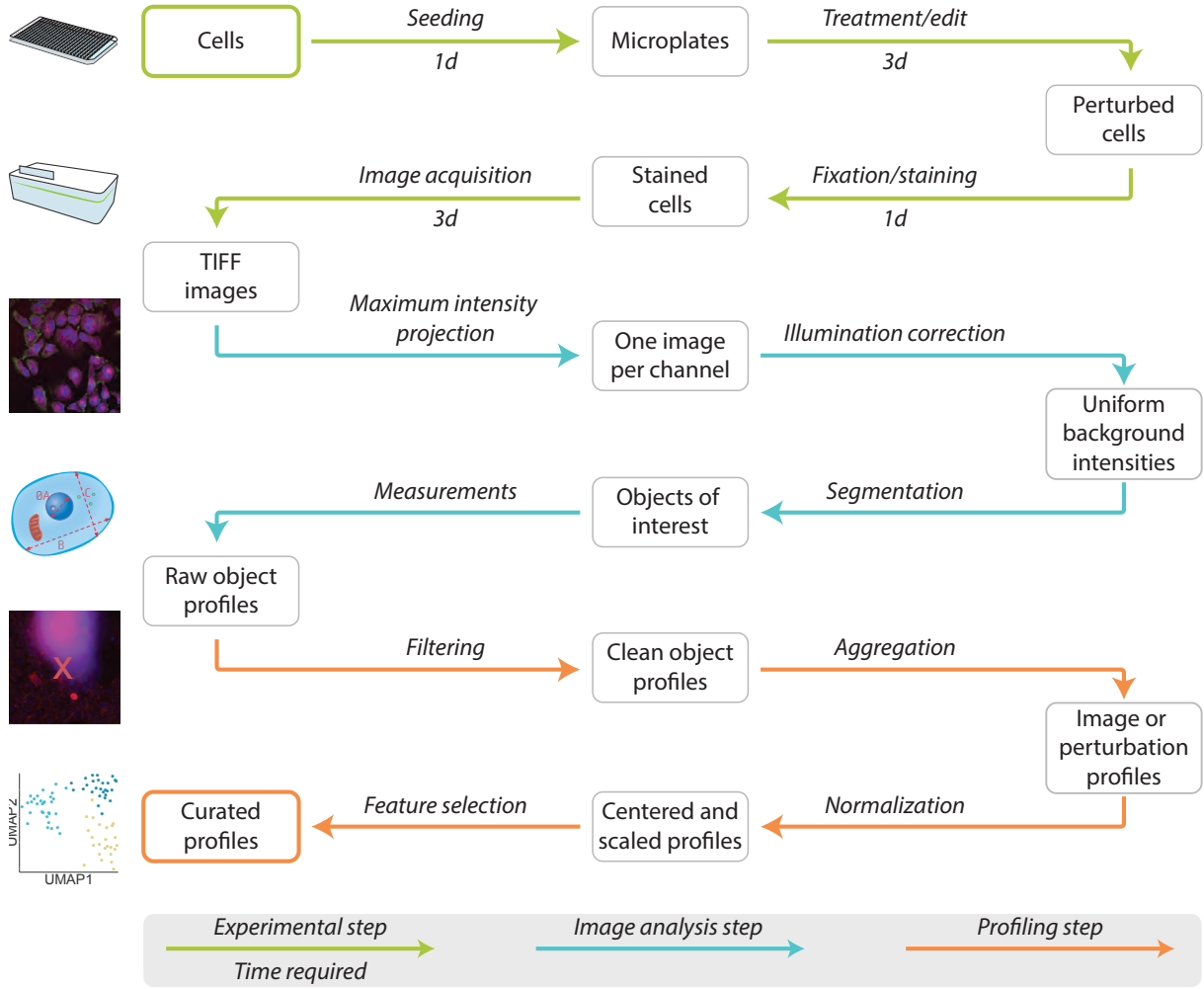


Figure 1.3: Common steps in high-content imaging experiments. For experimental steps, a typical duration is provided but this may vary greatly between experiments. The tasks are divided in three groups, corresponding to experimental, image analysis and morphological profiling steps. They are described in subsections 1.2.2, 1.2.4 and 1.2.5, respectively.

turbations, in yeast, *Drosophila* and human cells (Breinig et al., 2015; Caldera et al., 2019; Fischer et al., 2015; Heigwer et al., 2018; Mattiazzi Usaj et al., 2020), as well as the study of different aspects of healthy cellular organization. While pooled screens suffice for many applications (Doench, 2017), arrayed genetic screens can be conducted and it has been shown recently that the approach could be extended to imaging CRISPR screens (de Groot et al., 2018; Kim et al., 2018; Strezoska et al., 2017; Tan & Martin, 2016). Arrayed genetic screens have been used at several occasions to characterize gene function and links to cell shape in multiple experimental conditions and genetic backgrounds, for instance highlighting determinants of embryonic stem cell identity, or prioritizing metabolic disease genes (Bai et al., 2011; Chia et al., 2010; de Wet et al., 2020; Jiao et al., 2019; Vizeacoumar et al., 2010). Large-scale imaging experiments were also successful in basic biological research, with notable efforts to systematically describe protein location (Kraus et al., 2017; Ouyang et al., 2019b; Stenström et al., 2020; Sullivan et al., 2018), observe cell cycle protein regulation (Billmann et al., 2016; Mahdessian et al., 2021) and

identify key actors of cell division (Neumann et al., 2010). Finally, work in the field of HCS also fueled further technological developments, and a recent study demonstrated how imaging-based features allow for high-throughput cell sorting based on morphology (Schraivogel et al., 2022).

1.2.4 Image analysis

From hundreds of kilobytes for a dozen of images in the 1960s (Prewitt & Mendelsohn, 1966), bioimaging datasets saw a ramp-up to tens of terabytes and millions of images, or eight more orders of magnitude in storage space (Williams et al., 2017). Analysing HCI results thus requires efficient, dedicated computational tools. To this end, a large variety of software is available, as part of both commercial and open-source projects. Overall, guidelines for such analyses have been established and the main challenges to address are well characterized (Bougen-Zhukov et al., 2017; Caicedo et al., 2017). The flow of a typical analysis is represented in Figure 1.3.

Many steps are usually required to go from the raw microscopy images to the answers to the biological questions addressed in HCI experiments. The necessary conversion from raw images to quantitative measurements fits within the scope of bioimage analysis, which is a thriving interdisciplinary research field of its own (Meijering et al., 2016). Most methods used to process images and extract measurements are derived from the previous work done on HCI but also on other imaging modalities and in different study models, for which many analyses steps are shared. To streamline the image analysis process, several tools aim to combine multiple steps and offer visualization features. Some are widely adopted, such as ImageJ and its main distribution Fiji, which offer a lot of flexibility via custom plugins (Schindelin et al., 2012; Schneider et al., 2012), or CellProfiler, which is designed to be applicable to large screening datasets (McQuin et al., 2018). Some, like napari (Sofroniew et al., 2021), are general-purpose visualization toolkits, while others are dedicated to some specific imaging methods, like QuPath and its focus on digital pathology imaging (Bankhead et al., 2017). Other frameworks, such as ImJoy and ZeroCostDL4Mic, have a focus on simplifying the training of deep learning models and making the corresponding analyses more accessible and reproducible (Ouyang et al., 2019a; von Chamier et al., 2021). We now describe the main steps that need to be performed to analyse HCI data, whether using dedicated standalone tools or integrated in an end-to-end analysis framework.

If several focal planes were imaged, as is common with confocal microscopes, the different images may be stored in separate files, in a multi-channel TIFF file, or projected onto a single image. Maximum intensity projection is the most common method to do that, and consists in taking the maximal intensity value across the image stack for each pixel. In all cases, the images need to be processed to account for technical biases, such as aberrations in the optical path, non-uniform light source or shading, varying at different positions within the well and the plate (Coster et al., 2014; Singh et al., 2014). In some HCI experiments, empty wells may be imaged to serve as a model for background intensity distribution. Alternatively, retrospective methods may be used

on the noisy images: To correct uneven illuminations, a reference intensity distribution function can be computed either for each image individually, per position in the plate, or across a whole experiment. Different models have been proposed, accounting for the different steps separating the fluorescence emission from its detection by the camera. These models typically describe a foreground intensity F representing the actual dye fluorescence to be estimated as a function of the measured intensity values I , a background intensity B due to technical biases, and a shading factor S . In particular, an approach proposed by Coster and colleagues considers that one can model the signal as $I = S(F + B)$ and suggests an algorithm to remove first the effect of S then of B (Coster et al., 2014). In parallel, Singh and colleagues suggested a way to remove the shading factor while assuming the remaining background effect B to be sufficiently uniform not to impact the analyses (Singh et al., 2014). More approaches were proposed such as CIDRE, which formalizes illumination correction as an optimization problem to simultaneously account for S and B (Smith et al., 2015). Methods such as BaSiC are also accounting for temporal drift in timelapse experiments, which would otherwise violate assumptions of most other techniques, as imaging parameters may change over time and as the foreground regions in successive images are highly correlated (Peng et al., 2017). In practice, good strategies for illumination correction highly depend on the dataset and experimental setup, and exploring multiple approaches is recommended.

After preprocessing the images, the position of biological objects of interest needs to be determined in a task called instance segmentation. This means being able to attribute each pixel in the image to a specific object, such as a cell or nucleus, or to none. Traditional approaches rely on (i) setting an intensity threshold to first split the data between background and foreground, (ii) grouping contiguous foreground regions into separate objects, and possibly (iii) further splitting large objects, as they might actually represent two distinct objects in contact. Many algorithms were implemented to address each of these tasks. First, thresholding can be as simple as setting manually a global value which results in a good separation between foreground and background, or automated and more complex, for instance relying on intensity histograms or clustering-based algorithms (Sezgin & Sankur, 2004). Local methods also provide adaptive thresholds, which can be useful when the intensities are uneven across the image despite illumination correction. The second task is usually straightforward but might need to allow some degree of fuzziness to smoothen the objects and fill potential gaps. Third, different methods allow the borders of adjacent cells to be delineated based on shape or intensity values. The ‘watershed’ algorithm can be used for declumping cells, starting a diffusion process from local intensity maxima, expanding the corresponding regions until they meet and using this interface as the separation line between adjacent cells (Malpica et al., 1998; Vincent & Soille, 1991). Alternatively, if the shape of the cell is likely to provide more information about how to divide cell clusters, the watershed approach can be applied to distance-transformed images, to disregard original intensity values and focus on distances to local maxima. Other propagation methods are also suitable for a larger range of use cases (Carpenter et al., 2006). By implementing multiple methods for each task, CellProfiler can segment different object types for multiple imaging modalities (McQuin

et al., 2018). The approach can also be adapted to make the best use of specific aspects of the data and of prior biological knowledge available about the sample. For instance, dedicated methods are available when three-dimensional image stacks are acquired (Stringer et al., 2020), or can use previously segmented nuclei to help with the segmentation of the corresponding cell bodies (Jones et al., 2005). As many other fields, image analysis has been impacted by the advances of deep learning. The use of convolutional neural networks for segmenting cell images has significantly improved accuracy in several example tasks (Falk et al., 2019; Greenwald et al., 2021; von Chamier et al., 2021). However, the adoption of such methods by the bioimage analysis community at large was impeded by a difficulty to adapt these models to one’s individual experimental setup. As microscopes, imaging protocols and cell lines vary, the imaging data and their internal structure vary, and models that excelled in one context might underperform in another. Additionally, retraining large neural networks requires significant computational resources and a large annotated dataset, which is costly to produce. This situation can however be improved, and Cellpose, a generalist deep learning model based on intensity gradients, has recently gained traction, as it is flexible and performs well for a variety of applications out of the box (Stringer et al., 2020).

Once objects and regions of interest have been identified, it is possible to perform different types of measurements, thus quantifying biological properties across conditions. The traditional approach, coined feature extraction, relies on handcrafted geometrical metrics (Bougen-Zhukov et al., 2017; Caicedo et al., 2017). These features can refer to the size or shape of the cells, with metrics such as perimeter, area, eccentricity or form factor. It is also possible to measure the distribution of the intensity values for each dye, including summary statistics such as variance or median intensity per object, and more complex measurements, including textural descriptors such as Haralick features and Zernike moments (Pau et al., 2010). Moreover, some features can help summarize the local context of an object, for instance by counting neighboring cells or quantifying spot colocalization. Recently, new techniques emerged, powered by deep learning (Caicedo et al., 2018; Kensert et al., 2019; Lu et al., 2019). These approaches typically involve an intermediate ‘mock’ task used to guide the learning of a deep convolutional network, after which the weights in an intermediate layer of the model are used as features. Yet, learning a meaningful and compact quantitative profile directly from the data, a concept known as representation learning, is still a major open challenge (Bengio et al., 2012).

1.2.5 Morphological profiling

Most screens aim to identify hit perturbations leading to a phenotype of interest defined beforehand based on known mechanisms. For instance, it is common to track the expression of a fluorescent marker, or geometrical changes of the nucleus or cytoskeleton. Whereas such a targeted approach is efficient when investigating processes with a well-studied molecular basis, it is also possible to assemble data-driven quantitative representations of the induced

effects that integrate all the information available from the screen readout (Chandrasekaran et al., 2020). To do so, an exhaustive feature extraction step can be designed, where as many morphological descriptors as possible are measured, or such descriptive features can be derived directly from the images using representation learning techniques. The measured features can be assembled into vectors describing the cellular state observed in each condition, commonly called phenotypic or morphological profiles. These profiles have the potential to account for unexpected molecular phenomena and guide hypothesis generation, as they include little prior expectations of what changes might occur. However, these raw morphological profiles need to be curated to correctly depict differences between experimental perturbations and provide biological insight. To this end, Bougen-Zhukov and colleagues as well as Caicedo and colleagues published two introductory literature reviews in 2017, covering the available software solutions and providing guiding principles (Bougen-Zhukov et al., 2017; Caicedo et al., 2017). The main steps of morphological profiling are also summarized in Figure 1.3.

First, it is important to ensure that the profiles adequately describe the biological objects. While quality checks can be implemented at each analysis step, having access to the morphological profiles eases the identification of artifacts. An object with an unusually high intensity might correspond to a precipitation, and a cell with an extremely large area might be incorrectly segmented or declumped. A set of filters can be put into place to remove such erroneous data points. Compared to other experimental approaches, we here benefit from the fact that the images can often be directly interpreted through visualization, which greatly helps diagnose potential problems with the data.

Once multiple features describing the morphology displayed by cells for each perturbation have been obtained, one is left with the following matrix:

$$C = \begin{pmatrix} c_{1,1} & \dots & c_{1,p} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \dots & c_{n,p} \end{pmatrix}$$

where $c_{i,j}$ is the value of the j -th feature for the i -th cell, for n cells and p features. It can, however, be useful to aggregate the morphological features of all cells per image or per perturbation for several reasons. First and foremost, this has the advantage of mitigating the impact of extreme values present in individual measurements, thus the aggregated profiles may be more representative of the typical cellular state. By reducing the number of profiles, it also speeds up considerably all downstream analyses. Moreover, measurements are typically made on the basis of individual cells but the granularity may sometimes vary. By obtaining an aggregated value, it becomes possible to integrate cellular information with image information such as cell confluency in a given field of view, or with other information available about the perturbations, which could even arise from other experimental setups. Several options can be used to aggregate cell-level features, separately or in combination. Any real-valued variadic function f can be used

to define an aggregated profile matrix S as follows:

$$S = \begin{pmatrix} s_{1,1} & \dots & s_{1,p} \\ \vdots & \ddots & \vdots \\ s_{m,1} & \dots & s_{m,p} \end{pmatrix}$$

with

$$s_{i,j} = f(c_{k,j} \mid k \in A_i)$$

where A_i is the set of indices of the cells aggregated in the i -th profiles, out of m aggregated profiles. Most commonly, the mean and standard deviation are used as aggregation operators. It has also been shown that using higher-order descriptors of heterogeneity helps to discriminate between cellular perturbations (Rohban et al., 2019), although this approach is not widely adopted yet.

As hinted above, morphological features can describe a wide variety of cellular parameters. Each one can correspond to a different magnitude, ranging from normalized intensity values laying between zero and one, and cell area in pixels, which can reach multiple thousands. They can even be completely different types of data, such as discrete values for a number of neighbouring cells, pseudo-continuous values for the coefficient of overlap between two stains, or compositional data for the distribution of a dye intensity in a finite number of spatial bins. It is therefore essential to transform the data in order to give an equal importance to all variables, which is typically done by approximating a multivariate normal distribution in a high-dimensional feature space. This usually starts with the following transformation, which reduces the skewness of the data:

$$\hat{s}_{i,j} = \log(1 + s_{i,j} - \min(s_{1,j}, \dots, s_{m,j}))$$

with $\hat{s}_{i,j}$ denoting the corrected value of the j -th feature for the i -th profile. The data can then be centered and scaled to ensure the typical values of each feature are of a similar magnitude. The following normalization transformation can be used:

$$\hat{s}_{i,j} = \frac{s_{i,j} - \text{mean}(s_{1,j}, \dots, s_{m,j})}{\text{sd}(s_{1,j}, \dots, s_{m,j})}$$

The resulting features can then be considered to be Z-scores, as they represent how many standard deviations away from the mean feature value each profile is. For such transformations, the median and median absolute deviation (MAD) have also been used as robust counterparts of the mean and standard deviation (Caicedo et al., 2017; Heigwer et al., 2018; Malo et al., 2006). Batch and plate layout effects often cause systematic errors in HCS experiments, in part due to increased medium evaporation at the edges of the plate. This can be handled after normalization with methods such as Tukey’s median polish (Frederick & Tukey, 1977). If the layout of the negative controls permits it, it might also be possible to perform these corrections during the transformation step. In this case, each profile is scaled and centered on the control values in the same row or column of the plate only, instead of the full dataset. This also has the advantage

of giving a value that is indicative of the variation compared to the negative controls and not to all tested conditions, which might include a large portion of perturbations affecting cellular morphology.

After these transformations, it is possible to compare features with each other. In a typical HCS experiment, only features displaying more variation for some perturbations than among negative controls are of interest. Highly-correlated features are also redundant and can obscure more interesting structures in the data for downstream applications. A step of feature selection is therefore necessary. Several metrics can be used to quantify variation in the whole experiment and among negative controls such as variance, MAD or coefficient of variation. A simple method to decorrelate features is to proceed sequentially: Features can be ordered according to a property to prioritize, such as total variance, or randomly. The first feature is selected and everything correlated above a certain threshold for a chosen metric is discarded. The next non-discarded feature is then selected and the process is repeated until all features are either selected or removed (Breinig et al., 2015). Another approach consists in assembling a network where nodes represent features and links represent correlation above a given threshold. The network is then rendered disconnected using a minimal cut approach, meaning that no remaining features are correlated above the chosen threshold (Caldera et al., 2019). Depending on the setup and goal of the experiment, features might also be selected in other ways, for instance to favor interpretability, to match features identified as informative in the literature or to favor their reproducibility between replicates (Heigwer et al., 2018).

If the number of features left after the selection is too large for visualization or performing the downstream analyses at scale, dimensionality reduction can be applied. This is routinely done via techniques such as principal component analysis (PCA) or uniform manifold approximation and projection (UMAP) (McInnes et al., 2018). These methods have different drawbacks and advantages. For instance UMAP is able to account for non-linear relations between features while the axes of the PCA-reduced space can be interpreted as linear combinations of its input. However, all these methods are bound to distort the data in real-life applications as the information is compressed to a more compact representation (Cooley et al., 2020). In this thesis and the compiled articles, we call this curated reduced space the ‘morphological space’.

1.2.6 Distance between profiles

In most screens, we need to identify which perturbations are morphologically active, that is, that induce statistically significant changes in morphological descriptors. Morphological changes are often visualized using variations of radar charts (Breinig et al., 2015; de Wet et al., 2020), or by the schematic representation of an average cell for a given condition (de Groot et al., 2018; Khawatmi et al., 2021; Saïlem et al., 2015). This has the potential to conveniently portray the main characteristics of a perturbation and to quickly allow visual comparisons. We did so ourselves in multiple figures of the paper we introduce in subsection 3.2. However, a

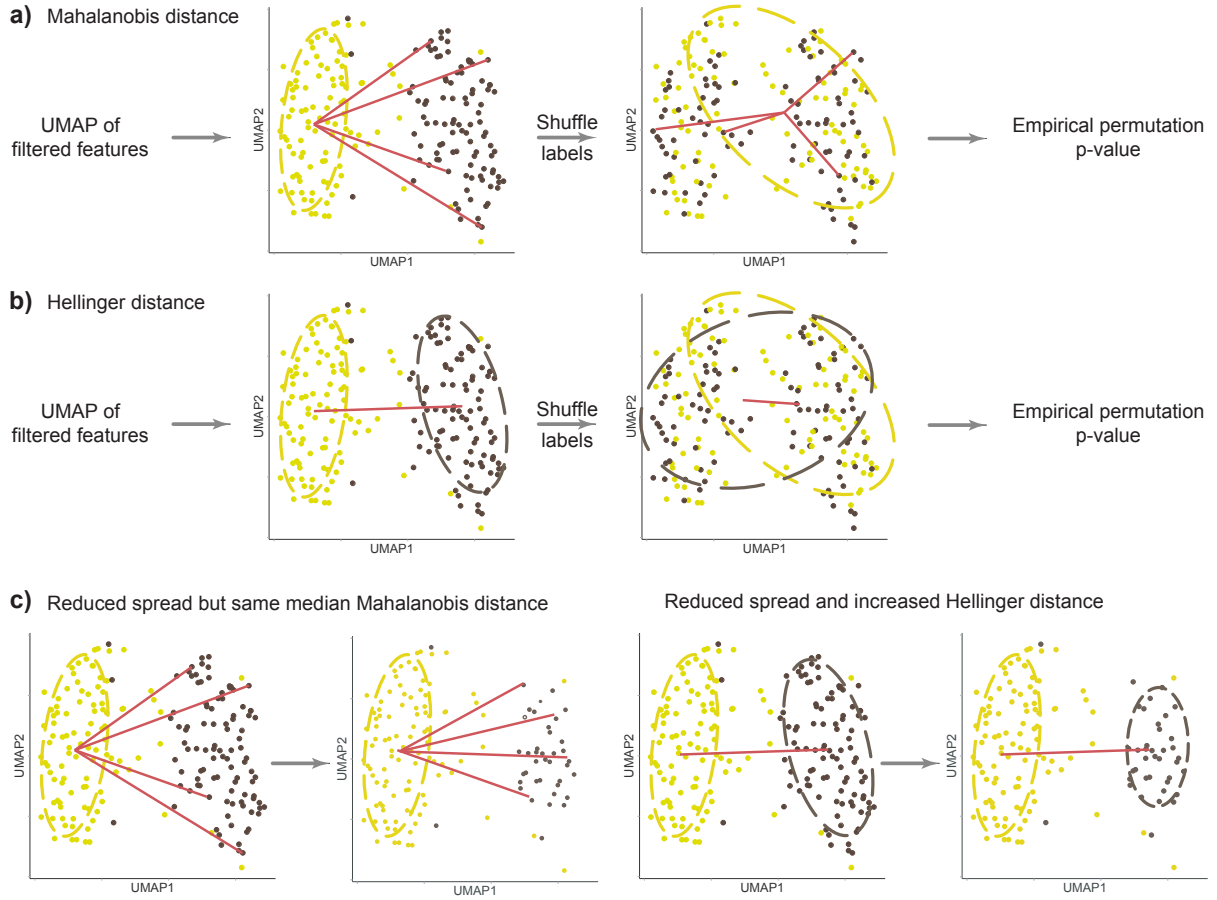


Figure 1.4: Distances accounting for dispersion to compare two statistical distributions. Both the Mahalanobis distance and the Hellinger distance can be used to quantify the difference between two populations but they differ in the way they account for dispersion. **a-b** Ways to obtain a value quantifying the statistical significance of morphological changes of perturbation profiles compared to reference profiles using the Mahalanobis or the Hellinger distance, respectively. **c** Impact of a reduction in the dispersion of the perturbation profiles for both distances. Red lines represent distances. Green points represent control profiles and brown points represent perturbation profiles. Ellipses symbolize confidence estimates for the covariance matrices used in the computation of the distances.

limitation of such graphs resides in that they do not convey a sense of sample size and statistical significance of the corresponding morphological changes. In addition to these plots, it is therefore essential to assess whether the changes observed are outside of the range of random variations of the measured parameters and can be interpreted with confidence. Conceptually, this can be seen as quantifying the likelihood of observing the distance measured between profiles in the morphological space.

Different methods can be used, depending on the properties of the data to be tested, such as perturbation reproducibility or strength. A useful paradigm was introduced by Hutz and colleagues, proposing the use of Mahalanobis distances combined with a permutation test to quantify the significance of morphological changes (Hutz et al., 2013). In brief, the procedure starts with a PCA of the data, scaling each axis by how much of the variance it explains. Then, one determines the center μ_R and covariance Σ_R of reference profiles, such as the negative

controls, as well as the center μ_P and covariance Σ_P of a perturbation's profiles. These variables are used to obtain a metric D_M analogous to the Mahalanobis distance, and computed as:

$$(D_M)^2 = (\mu_P - \mu_R)^T \hat{\Sigma}^{-1} (\mu_P - \mu_R)$$

where $\hat{\Sigma}$ is the cardinality-weighted mean of Σ_R and Σ_P . To contextualize the value of D_M , one empirically estimates its distribution under the null hypothesis that there is no difference between reference and perturbation by shuffling the labels of the reference and perturbation points. The resulting empirical p-value is coined the mp-value, and used to define which perturbations are morphologically active.

The Mahalanobis distance has the advantage of taking the dispersion of the data into account, as opposed to the Euclidean distance, which gives the same weight to all dimensions. One limitation, sketched in Figure 1.4, is that it only provides a distance between a point and a distribution. In the mp-value method, this is mitigated by using $\hat{\Sigma}$ instead of Σ_R as in the standard definition of the Mahalanobis distance. Other distances might be more suited to morphological profiles, as they are symmetrical and intended to account for the dispersion of both sets of profiles. In particular, the Hellinger distance D_H is defined for multivariate normal distributions as:

$$(D_H)^2 = 1 - \frac{\det(\Sigma_P)^{\frac{1}{4}} \times \det(\Sigma_R)^{\frac{1}{4}}}{\det(\Sigma')^{\frac{1}{2}}} e^{-\frac{1}{8}(\mu_P - \mu_R)^T (\Sigma')^{-1} (\mu_P - \mu_R)}$$

with $\Sigma' = \frac{\Sigma_R + \Sigma_P}{2}$. This distance satisfies both criteria. However, the fact that D_H is influenced by both Σ_R and Σ_P simultaneously also means that it can be tedious to interpret distance changes from this value alone, as they might be explained by both effects on location and dispersion.

Another challenge when analysing HCI experiments is that technical artifacts are common, leading to outlier profiles. Some types of statistics are better suited to dealing with extreme values and outliers. This property is called robustness and can be quantified with metrics such as the breakdown value, which is the minimal fraction of outliers in a population that is sufficient to make the estimator “arbitrarily bad” (Hubert & Debruyne, 2009). Previous studies demonstrated the benefits of using robust statistics to deal with the noise and heterogeneity of HCS data (Caicedo et al., 2017; Heigwer et al., 2018; Malo et al., 2006). This approach was mainly constrained to the normalization steps, and we extended it to the downstream analysis steps as well, including the quantification of distance between morphological profiles. This focus on robust statistics, together with the use of the Hellinger distance, formed the conceptual basis for the work described in the article presented in section 3.1.

1.3 Network-based modelling of biological systems

Experimental results are best interpreted when taking into account what is already known about the system under study. For instance, Neumann and colleagues chose to classify cellular morphology in groups based on cell cycle and nuclear state rather than describing morphological changes with no prior expectations. This prior knowledge guided their analysis and allowed them to identify molecular regulators of mitosis (Neumann et al., 2010). Several studies also showed that constraining the architecture of deep learning models based on known biological properties can be beneficial, for instance when trying to estimate signalling pathway activation or patient survival. This approach led to a good performance of the models, and informed about the molecular mechanisms underlying their predictions at the same time (Elmarakeby et al., 2021; Fortelny & Bock, 2020; Ma et al., 2018). Similarly, it is essential to identify ways in which morphological profiles, as introduced in the previous section, can be contextualized, and how the information they contain can contribute to our molecular understanding of cells and their response to perturbations. In recent years, a large body of work has been dedicated to so-called systems biology approaches, aiming to bring together scattered molecular insight from multiples sources into a single comprehensive representation. Systems biology provides both a conceptual framework and methods for the integration and contextualization of heterogeneous data, which may also be leveraged for HCS analyses.

This section provides key concepts in systems biology and network medicine. More specifically, we start with a brief introduction to the necessity of studying living organisms as complex systems made of many interacting elements, which is made possible by high-throughput experimental methods. In the second subsection, we perform a literature review presenting networks as a suitable representation of such systems. We observe how different biological processes correspond to networks at varying scales, ranging from molecules to populations, and how they are, taken together, underlying health and disease. In a last subsection, we expand on network-based approaches aiming to better understand and treat diseases, referred to as network medicine. In particular, we focus on how network medicine provides a framework for the integration of heterogeneous data from high-throughput experiments, potentially generated via HCS. Finally, we conclude by considering the promises of network medicine regarding clinical translation.

1.3.1 Studying biological systems as a whole

All living organisms share a collection of unusual properties, rendered possible by the interplay between a limited number of basic building blocks. For instance, DNA is made of only four types of nucleotides, and most biological macromolecules fall within four classes: carbohydrates, lipids, nucleic acids and proteins. Such simple elements can interact in an astonishingly large

number of ways, due to the mathematical concept of combinatorial explosion (Schuster, 2000). In this way, there are already more possible nucleotide sequences of length 140, which is shorter than a typical Illumina sequencing read, than atoms in the universe. The traditional reductionist approach to molecular biology attempts to describe the roles of genes and molecules in isolation and to proceed sequentially to progressively describe complete organisms. Yet, there are too many possible combinations of the basic components forming living systems to be studied exhaustively. Thus, the reductionist approach is limited in its ability to retrieve the emergent properties of cells and organisms (Lazebnik, 2002). The field of systems biology explores ways to integrate all the findings about the components of a biological system (Stéphanou et al., 2018). First, systems biology offers a conceptual framework to think about biological information (Ideker et al., 2001). The results from high-throughput perturbation experiments conducted in multiple studies are combined into models of the interactions and pathways that explain global changes. Second, systems biology provides concrete computational tools, including databases and methods to navigate the complexity of the new models, which can no longer be easily pictured and visualized as simple sequential and linear processes. For instance, the Reactome database lists thousands of actors involved in thousands of reactions within signalling and metabolic pathways (Fabregat et al., 2016). Its curators split all these complex relations in hundreds of smaller functional units at different levels of granularity following a hierarchical structure, so that each term can be visually inspected separately. Systems biology has branched into many derivatives, such as systems genetics and systems medicine, in which the emphasis is not given anymore on single causal genes. On the contrary, studies explore how the entire genotype and different modes of interaction can lead to emerging properties, and dictate the resulting phenotype and disease status (Goh et al., 2007; Menche et al., 2015). Such studies often aim at uncovering general rules, which make it possible to translate the knowledge accumulated to other systems or to adapt to the characteristics of different patients more precisely (Apweiler et al., 2018).

Although considering biological systems as a whole is beneficial in most applications, some research directions are more amenable to a systems approach or more prone to consider their study object as a complex system than others. The nervous system is a rare case of a complex cellular organization structured as a tangible physical network, where information is propagated through neurons in the form of electrical and chemical signals. This makes systemic studies appear as a natural choice, and the complete wiring diagram of the nervous system of *Caenorhabditis elegans* was already mapped in 1986 (White et al., 1986). However, scaling this approach to humans is challenging in part due to major differences in scale and connectivity patterns. Methodological changes might help to address these challenges in the future (Jonas & Kording, 2017). Another field with a clear connection to systems approaches is immunology, where the interaction and cooperation between different cell types is key to a functional immune system (Rieckmann et al., 2017). This can be seen as an heterogeneous system, comprised of highly-specialized cell types which must communicate through molecular signalling, in ways analogous to other complex systems such as human social networks (Bergthaler & Menche,

2017).

Adopting a system-centered view of disease etiology led to notable changes to research strategies. As opposed to Mendelian diseases, in which a handful of inherited DNA variants can lead to a disease, complex diseases might be influenced by a larger number of genetic elements simultaneously (Botstein & Risch, 2003). For instance, dozens of variants may participate mechanistically in the trait of interest, affect large protein complexes or pathways indirectly associated with the trait, or act through transcription regulation. In this way, many genes might be separately identified as related to the trait, yet none would explain the presence or absence of the trait on its own. Even compiling the hundreds or thousands of significant genetic findings in large genome-wide association studies often does not suffice to explain the prevalence of transmissible traits. This is often referred to as the missing heritability problem (Manolio et al., 2009). The omnigenic model of complex traits suggests that some traits can be explained by a large number of variants with weak effects spread across open chromatin regions, of which only a limited number of core genes are directly and mechanistically causative (Boyle et al., 2017). The scale and high density of the gene regulatory network would thus explain parts of the spread of trait heritability between many loci, as many regulatory variants may indirectly affect one of the core genes. However, many observations also do not concur and better align with a completely polygenic model, without a clear separation between core and peripheral genes (Wray et al., 2018).

Finally, the emerging tendency to consider that cell properties and behaviors emerge from the interconnection of subsystems at different scales can be seen as another important paradigm change. This can be achieved by departing from considering the gene as the sole biological unit meaningful to explain phenotypical changes, and by realizing that different biological structures at different scales are important when studying different processes (Ma et al., 2018; Qin et al., 2021). Individual genetic mutations are sometimes enough to directly disrupt an essential molecular interaction (Meyer et al., 2018), while in other cases a phenotypic change might be the result of altered cell-cell interaction patterns (Vladimer et al., 2017).

1.3.2 Complex Networks in Health and Disease

Loan Vulliard and Jörg Menche. Systems Medicine. Available online: August 28, 2020.

DOI: <https://doi.org/10.1016/B978-0-12-801238-3.11640-X>.

Complex networks are at the core of the study of biological systems. They are mathematical graphs representing components as nodes and interactions as edges between nodes, and they typically exhibit some special structural properties (Barabási & Oltvai, 2004). Networks are omnipresent in modern biology. Network science suggests that the interplay between the components of a complex system can explain the emergence of general properties of this system. For instance, gene expression is regulated by multiple layers of signalling and compartmentalized physical interactions involving DNA, RNA, proteins and metabolites in the cell. These processes can be modelled by gene regulatory and interaction networks. Networks are also common in biomedical applications, where they permit the characterization of shared properties of disease-related genes, which can potentially guide clinical diagnostic and treatment (Barabási et al., 2011). In the following book chapter, we describe how prevalent and useful biological networks are. We cover the most common types of networks encountered in biology and which methods are available to interpret them. We also emphasize that understanding how networks relate across scales is the key to understand drivers of health and disease.

This invited review was published as part of the collection “Systems Medicine: Integrative, Qualitative and Computational Approaches” edited by Prof. Olaf Wolkenhauer. The aim of this collection was to compile current state-of-the-art knowledge in the field of systems medicine. Two entries were provided by our research group. The first one reproduced below, is written by Prof. Jörg Menche and myself and introduces basic concepts for navigating the field’s literature. The second contribution covers more in depth the measures which can be used to characterize properties of complex networks at different scales, from local structures to global topology (Hancock & Menche, 2021).

Complex Networks in Health and Disease

Loan Vulliard and Jörg Menche, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

© 2020 Elsevier Inc. All rights reserved.

Introduction	1
From Network Structure to Biological Function	2
Random Networks as Reference Models	2
Molecular Interaction Networks	3
Beyond Physical Interactions	4
From Molecules to Organisms	4
Global Networks in Epidemiology	5
Summary and Outlook	5
References	6

Glossary

***ome** Suffix that is commonly used to emphasize that a collection of individual elements is considered as a whole. The entirety of all genes, for example, forms the genome. Compare with interactome and exposome.

Complex network Complex networks provide holistic models of systems that consist of many interacting elements. Each individual element is represented as a node (also called 'vertex') and each interaction between two elements as a link (or 'edge'). The connection patterns within complex networks are typically neither completely regular, nor completely random.

Disease module Genes associated with a particular disease are not scattered randomly within molecular networks, but aggregate in certain neighborhoods or 'disease modules.'

Exposome Organisms are constantly exposed to a multitude of biological and chemical factors through their environment, collectively referred to as 'exposome.' The dynamic combination of the internal biological state (homeostasis) and external influx (exposome) results in health or disease.

Functional networks Networks in which links represent indirect relationships, for example correlated activity patterns in functional brain region networks or shared biological processes in gene similarity networks. Compare with physical networks.

Herd effect Phenomenon observed in epidemiology that the immunization of one part of a population also decreases the risk of infection for the other part that is not immune. The herd effect is important to protect sub-populations that cannot be treated (e.g., immunocompromised people) and allows for the overall success of immunization campaigns even under partial compliance.

Interactome In analogy to the genome, the interactome represents the complex network of all molecular interactions within a biological system. More specifically, the term often refers to physical protein-protein interactions within the cell. Over the last two decades the interactomes of several organisms have been mapped out systematically, ranging from model organisms to human.

P4 medicine Predictive, preventive, personalized and participative medicine. This approach aims to improve healthcare by progressively reducing the need for one-size-fits-all palliative treatments by leveraging modern therapeutic solutions and personal patient data such as genome or microbiome sequencing.

Physical networks Networks in which links represent physical interactions, such as physical binding between proteins in interactome networks, or synapses in neural networks. Compare with functional networks.

Protein-Protein Interaction (PPI) Cellular processes rely on the coordinated interaction between different proteins. Several experimental methods are available to detect such interactions, for example yeast two-hybrid approaches or affinity purification combined with mass spectrometry.

Introduction

In the 19th century, brilliant pioneers like Mendel, Darwin and Semmelweis required little more than meticulous observation of their surroundings for developing their theories on inheritance, evolution and infectiology that would become the foundations of modern biology and medicine. Since then, a range of sophisticated technologies have been developed that now allow us to observe biological systems at molecular resolution. However, our rapidly growing knowledge at the molecular level also revealed the fundamental limitations of traditional reductionist approaches that aim to understand complex biological systems by dissecting their individual elements (Greene and Loscalzo, 2017; Stéphanou et al., 2018). It is becoming increasingly clear that many system-wide phenomena cannot be understood in this fashion, and that often the 'whole is more than the sum of its parts.' It is thus essential to systematically study not only the isolated elements of these systems, but also their interactions. These interactions are

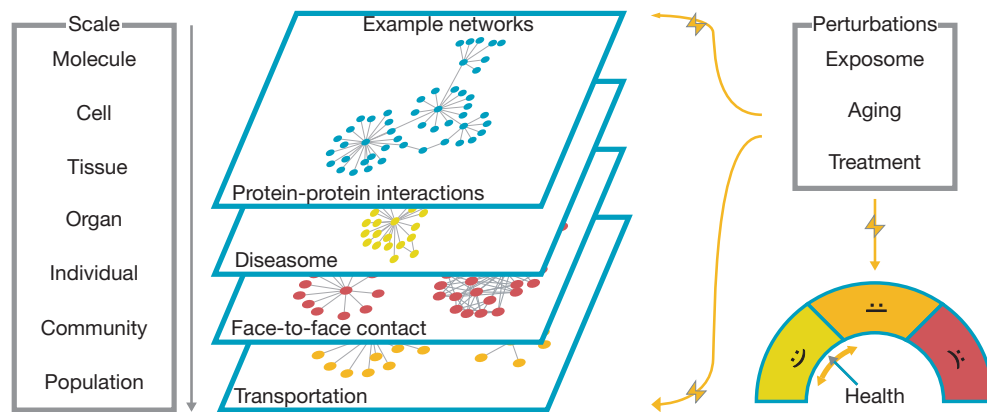


Fig. 1 The processes involved in human health and disease range from the molecular to the population scale. Networks provide a unifying framework for describing and investigating the complex interactions that occur both within and across scales: They can be used to describe the molecular interactions forming the basis of all biological processes within the cell, as well as the social interactions that form the basis for the spread of infectious diseases. Diseases can be conceptualized as perturbations of these intricate systems. These perturbations may be internal, for example genetic mutations, or external, for example environmental exposures.

key to understand the emergence of novel properties and behaviors, in particular when moving across different scales, i.e., from molecules to cells, tissues, organs and organisms or entire populations (Fig. 1). *Complex networks* provide a natural framework for systematically investigating the various relationships between the constituents of biological systems within and between scales.

In the following we aim to provide an overview of how tools and concepts from network theory may help address important fundamental and practical challenges in biology and medicine. We start by reviewing key relationships between structural network properties and functional characteristics of the represented biological system. We then introduce frequently used networks, from the molecular level of protein-protein interactions within the cell, all the way to the level of transportation networks that span the globe. We conclude by highlighting a few future challenges in this highly active and dynamic research field.

From Network Structure to Biological Function

Network theory provides a versatile and general toolbox for investigating complex systems composed of interacting elements. In the most generic case, each element of the system is represented by a node and each interaction between a pair of nodes by a link (Fig. 2). In mathematical terms, the collection of all nodes and links is also called a ‘graph.’ This simple definition can be extended, for example by adding weighted or directed links, by including time-dependence or different layers of connectivity between the nodes, resulting in so-called weighted, directed, temporal or multi-layer networks, respectively.

Networks can be characterized at different levels, ranging from the level of individual nodes (e.g., their number of connections, or their centrality within the network) to the level of groups of nodes (e.g., their connection density), to the global level of the entire network (e.g., the distribution of the number of connections per node across all nodes in the network). The finding that these properties can be associated with important biological characteristics makes network theory a valuable tool in biology and medicine. For example, proteins that are located at a highly central position within molecular interaction networks have been shown to perform important roles in the cell, whereas more peripheral proteins are often less essential (Piñero et al., 2016; Costanzo et al., 2019). Densely interconnected groups of nodes correspond to functionally closely related groups of proteins (Barabási et al., 2011). Similarly, genes that are associated to the same disease tend to aggregate in specific *disease modules* within molecular networks (Menché et al., 2015; Ghiassian et al., 2015).

Random Networks as Reference Models

In order to assess the magnitude and statistical significance of an observed network characteristic, suitable random controls are needed. Network theory provides a wide range of well-studied random graph models that can be used as reference (Albert and Barabási, 2002; Piñero et al., 2016). The most basic model is the classic random graph, in which a given number of links is distributed randomly among a given number of nodes (Erdős and Rényi, 1960). In contrast to many real world networks, including most molecular networks, random graphs do not contain highly connected nodes, so-called hubs, underlying their importance as they could not have emerged by chance alone. More advanced reference models can be introduced by keeping additional properties of the original network constant during the randomization procedure. Rewiring algorithms, for example, keep the number of links per node fixed and have been used to uncover basic design principles of gene regulatory networks, such as the tendency of highly

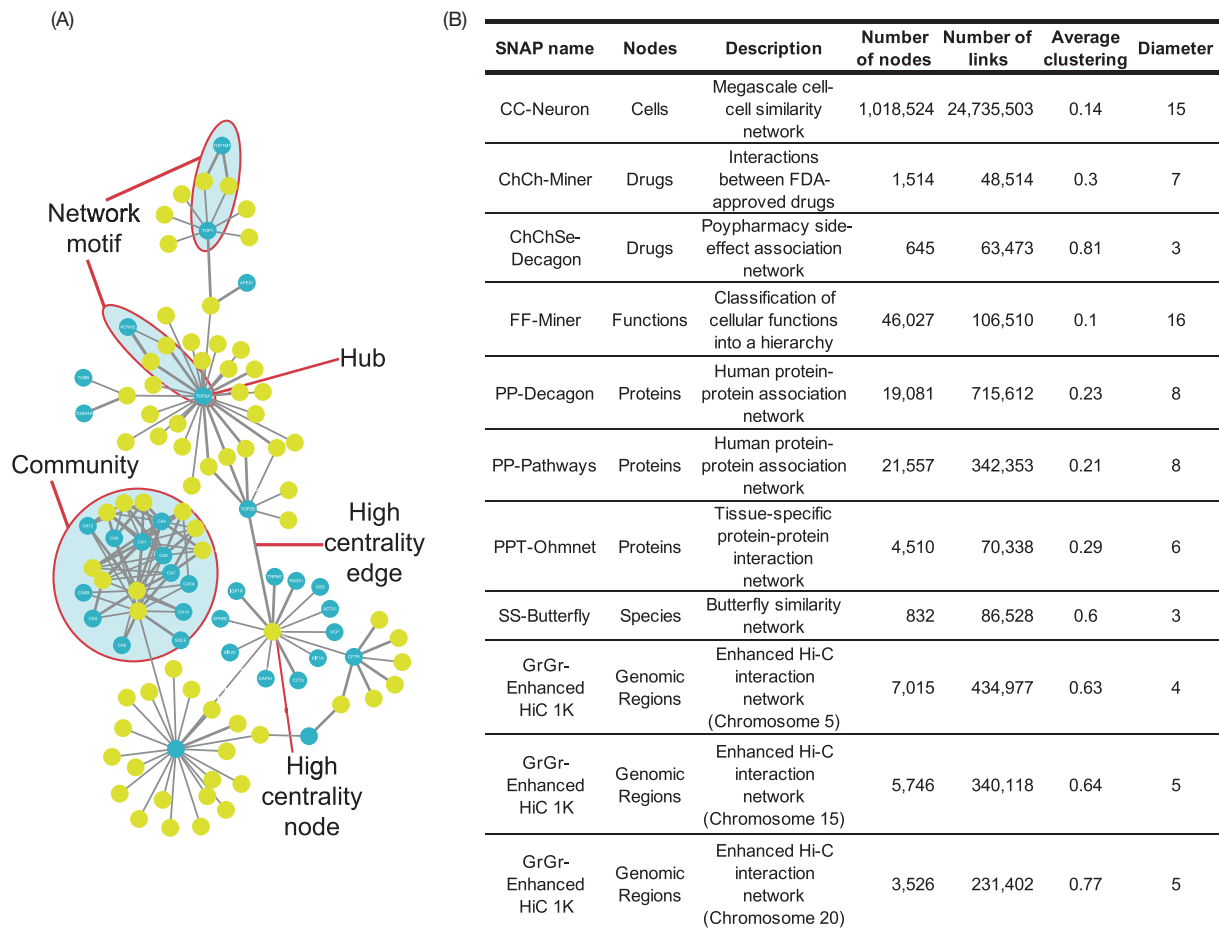


Fig. 2 (A) Illustration of basic network characteristics for a protein-chemical interaction network (data from the BioGRID database (Oughtred et al., 2019), obtained and laid out using the NDEx platform (Pratt et al., 2015)). Green and blue nodes represent chemicals and proteins, respectively, that interact with each other. Several network characteristics are highlighted: network motifs are small recurrent connection patterns; hubs are nodes that have a large number of neighbors; densely interconnected groups of nodes are called 'communities' (also 'modules' or 'clusters'); high centrality indicates that the respective nodes/edges are of structural importance to the network. (B) Overview of different biological networks and their basic characteristics (data from the BIOSNAP repository (Leskovec and Sosić, 2016; Zitnik et al., 2018)). The average clustering coefficient is a measure for the overall local density of a network; the diameter of a network is given by the greatest distance between any pair of nodes.

connected hubs to avoid each other (Maslov and Sneppen, 2002) or the discovery of network motifs, i.e., recurrent connectivity patterns among small groups of nodes (Milo et al., 2002).

Also dynamic aspects of networks, such as their growth over time, can be assessed through random models. An important class of networks that are characterized by the presence of hubs are so-called scale-free networks (Albert and Barabási, 2002). These networks emerge by iteratively adding new nodes to the network, such that they have a tendency to form links with already highly connected nodes. The basic model can be extended by adding a latent fitness to each node (Bianconi and Barabási, 2001) to investigate the role of evolutionary processes and positive selection of genes with high fitness in shaping the structure of molecular networks.

Molecular Interaction Networks

The first layer of information transfer and transformation from genotype to phenotype is mediated by molecular networks within the cell. In analogy to the genome representing the blueprint for all molecular components, the collection of all their interactions is referred to as the *interactome*. The interactome thus represents the blueprint for the collective functions that emerge from interactions between individual components. Most commonly, the term interactome is used specifically for *protein-protein interaction (PPI)* networks. Over the last two decades, genome-scale PPI networks have become available for a variety of species (Alanis-Lobato et al., 2017; Oughtred et al., 2019). PPIs can be mapped out systematically using yeast two-hybrid approaches (Rolland et al., 2014) or

mass spectrometry based methods (Huttlin et al., 2017). The most comprehensive PPI networks also incorporate results compiled from numerous small-scale experiments from the literature (Oughtred et al., 2019) or computational predictions (Kovács et al., 2019).

PPIs are not the only molecular interactions within the cell that are biologically relevant and can be experimentally assessed. Other important molecular networks include metabolic networks and signaling cascades (Choudhary and Mann, 2010; Fabregat et al., 2016). Moreover, links may also represent indirect relations, for example in gene regulatory networks, where one gene can act on another via transcribed RNA or a translated protein, resulting from the binding of transcription factors and regulatory elements to the genetic material, which can be assessed experimentally through chromosome conformation capture techniques (Babaei et al., 2015).

Over the last two decades, numerous relationships have been uncovered between the structural characteristics of molecular networks and the function of the systems that they represent. In PPI networks, for example, connection patterns such as the number of interaction partners, network distance between proteins or densely interconnected network neighborhoods are directly related to biological functions in both healthy and disease states (Barabási et al., 2011; Caldera et al., 2017; Meyer et al., 2018). In a network context, disease states can often be identified with localized perturbations of the underlying molecular network. Such perturbations may be internal, for example genetic mutations associated with severe hereditary diseases (Köhler et al., 2008), or external, such as chemical or other environmental exposures (Kalia et al., 2019). Collectively, the set of all such external factors is called the *exposome*. Given the broad nature of this term, it is unclear whether a comprehensive mapping of the exposome and its impact on the interactome is at all achievable. First attempts in this direction focus on specific exposures, for example the impact of toxicants on metabolite networks (Kalia et al., 2019; Veneman et al., 2017). These approaches enable on the one hand the inference of which chemicals a system has encountered, and offer on the other hand an opportunity to elucidate the response mechanisms following a particular exposure.

Beyond Physical Interactions

The links in the molecular networks discussed above represent physical interactions that can be directly measured. In addition to these *physical networks*, we can also construct *functional networks*, where links represent more indirect relationships or similarities. The most commonly used functional networks are co-expression networks, where two genes are linked if their expression levels were found to be correlated across different experimental conditions (Saha et al., 2017). Other important examples are genetic interaction networks, where a link between two genes indicates that the phenotype of their combined knock-out deviates from the expectation based on the individual knockouts (Costanzo et al., 2016; Rauscher et al., 2018), drug-drug interaction networks, in which links connect non-additive drugs (Caldera et al., 2019) or chemical networks, in which compounds are linked based on structural similarity (Lo and Torres, 2016).

Functional networks may also contain various types of nodes, connecting for example genes and drugs: A genome-wide screen in *Saccharomyces cerevisiae* has recently been used to map out the interactions between 1377 chemical compounds and 177 genes (Piotrowski et al., 2017). Systematically exploring pairwise combinations of cellular perturbations has great potential for functionally annotating individual components, such as genes, drugs or environmental factors, as well as for identifying the involved molecular pathways and, more generally, for elucidating the fundamental rules that underlie the cellular response to combinations of perturbations.

Similarity networks were further used to study individual exposomes, by connecting co-occurring species or chemicals, revealing temporal and environmental patterns such as compounds that were released simultaneously during rainy days (Jiang et al., 2018). This exemplifies the potential to investigate different aspects of a biological concept (here the exposome) through complementary network approaches.

From Molecules to Organisms

While biological processes span a wide range from molecules to cells, tissues, organs and whole organisms, the networks at the molecular level are the most studied and best understood at this point. This reflects their importance as the primary interface between genotype and phenotype, but also the fact that they are more easily accessible experimentally compared to other relevant networks.

At the level of cellular organization, the neural networks that constitute the nervous system have probably received most attention (Bullmore and Sporns, 2009). Considerable efforts are made to systematically map out neural networks, in humans, as well as in model organisms. Similar to the different types of molecular networks introduced above, neural networks may also either represent direct cellular networks, where nerve cells are connected through synapses, or functional networks, in which regions of the brain are linked if they show correlated patterns of activity. The first complete direct neural network was resolved as early as 1986 for the worm *Caenorhabditis elegans* (White et al., 1986). For higher organisms, only partial maps are available to date, for example in mice (Briggman et al., 2011; Bock et al., 2011), but also in human (Glasser et al., 2016), if only at a very coarse grained level. Mapping out the complete human ‘connectome’ of all our brain cells will likely remain out of reach for many years due to its staggering size (Sporns, 2013).

A similarly complex and important biological system is the immune system, whose primary objective is to maintain the normal function of an organism under constant threat by internal and external challenges, ranging from tumor cells to viral infections. Given the diversity of participating organs, cell types and molecules, it has been proposed to conceptualize the immune system as a multi-layered network (Bergthaler and Menche, 2017; Rieckmann et al., 2017; Kveler et al., 2018). The nodes in this network represent cells, links represent communication through signaling molecules, such as cell-surface receptors or secreted molecules. Different layers may represent different contexts, such as organs, tissues or activation status.

Global Networks in Epidemiology

An effective response to a viral or bacterial infection is not only critical for individual organism, but may also be seen in the much larger, potentially world-wide, context of epidemics. To accurately model the spread of a contagious disease, we must understand both the social networks of personal interactions, as well as the local and global transportation networks along which people travel (Pastor-Satorras et al., 2015). Interestingly, it was shown that not only infectious diseases are transmitted across networks of social interactions, but also other sociological and health-related conditions, such as smoking behavior (Christakis and Fowler, 2008), weight gain (Christakis and Fowler, 2007) or happiness (Fowler and Christakis, 2009).

Mathematical models have a long history in epidemiology and date back to the early 20th century (Kermack and McKendrick, 1927). Classical models divide a population into three compartments, in which people are either susceptible to an infection (S), currently infected (I), or recovered (i.e., immunized) or otherwise removed (R) from the susceptible pool. The temporal dynamics of these SIR models can be described by differential equations. Early models typically assumed ‘uniform mixing,’ i.e., an equal probability for any infected individual to contaminate any susceptible individual. More recently, these models have been significantly improved by considering the relevant social and transportation networks that underly the disease spreading process (Wang et al., 2017). For instance, the structure of the face-to-face contact network has a profound impact on how fast and how far a contagious disease may spread among a population (Pastor-Satorras et al., 2015). Likewise, the efficiency of different immunization strategies can only be fully understood when taking these networks into account.

Smallpox is currently the only human infectious disease that was successfully eradicated through immunization. There are two key aspects for the success of an immunization campaign: First, the effective access to immunization, which includes the availability of a vaccine, but also the individual willingness to get vaccinated. The latter may decline along with the disease prevalence, even in countries where immunization is compulsory. It has been shown that a better understanding of the herd effect improves the adherence to such programs (Brockmann, 2017; Betsch et al., 2017). Second, the impact that an immunization of a certain subpopulation has on the spreading of the disease, i.e., how it reduces the contagion rate in total, as well as within smaller subcommunities. This aspect can be studied from a network theory point of view. The structure of the social face-to-face contact network determines whether bottlenecks (such as high centrality nodes or links) could prevent a disease outbreak and whether some communities are more at risk than others. This enables the design of more efficient quarantine strategies with maximal impact on the social network connectivity. In the past, epidemiological modeling was essential for example in handling the avian influenza outbreak in 2005 (Longini et al., 2005). More recently, great efforts have been made to profile and constrain the spread of the 2019-nCoV virus, including charting the phylogeny of viral samples (Hadfield et al., 2018) and mapping the spreading risk based on global transportation networks, which allow for predicting disease spread much more accurately than maps based on geographic distance (Brockmann and Helbing, 2013).

Summary and Outlook

Networks provide a powerful framework for investigating biological systems ranging from the molecular to the global scale. A key factor for the success of network theory in biomedical applications is that many structural network characteristics can be related to functional properties of the respective biological system. In molecular networks, for example, densely connected node communities often correspond to proteins involved in a particular cellular process. Likewise, disease associated processes can be identified with specific connectivity patterns between groups of perturbed nodes.

An important open question in this context is how exactly different network perturbations influence each other. For example, it has been found that a network overlap between a drug-induced perturbation and a disease associated perturbation may either indicate an effective treatment of the respective disease, but also the opposite, namely that the disease may be a side effect of the treatment (Cheng et al., 2019; Guney et al., 2016). This highlights an important methodological and conceptual limitation of current network approaches: We are still lacking a systematic understanding of the combined effect of independent perturbations, in particular when considering complex phenotypes. To date, most large-scale experimental efforts for elucidating the combined effect of perturbations relied on relatively simple, one-dimensional readouts, such as growth assays, for example in the characterization of genetic interactions (Kuzmin et al., 2018) and drug-gene interactions in yeast (Piotrowski et al., 2017). More recently, more informative readouts have been employed as well, such as high-content imaging or next-generation sequencing, which allow for a much more detailed assessment of the interactions between different perturbations. Using these high-dimensional readouts it is possible to identify different types of interactions between perturbations (positive, negative), as well as their direction. Furthermore, high-dimensional readouts allow for the identification of interactions that lead to the emergence of entirely new phenotypes. The

first studies aiming to map out such high-resolution ‘perturbome’ networks were based on morphological changes induced by combinations of genetic perturbations in a model organism (Fischer et al., 2015) and combinations of drug perturbations in cell lines (Caldera et al., 2019), respectively.

Another major focus of recent network-based biomedical research is the integration of the diverse data describing different levels of biological organization. While combinations of different ‘omics’ data, e.g., genomics, transcriptomics, proteomics, metabolomics and microbiome data, are becoming more and more common in basic research, their translation into clinical applications is still scarce (Karczewski and Snyder, 2018), despite their potential for applications in *P4 medicine* being widely recognized (Apweiler et al., 2018). Network approaches can offer valuable contributions to solving current technical and conceptual challenges in integrating multi-omics and multi-scale data (McGillivray et al., 2018). Indeed, concrete translational impact is the ultimate ambition of network biology and network medicine.

References

- Alanis-Lobato G, Andrade-Navarro MA, and Schaefer MH (2017) HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Research* 45(D1): D408–D414.
- Albert R and Barabási A-L (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74(1): 47–97.
- Apweiler R, Beissbarth T, Berthold MR, Blüthgen N, Burmeister Y, Dammann O, Deutsch A, Feuerhake F, Franke A, Hasenauer J, Hoffmann S, Höfer T, Jansen PL, Kaderali L, Klingmüller U, Koch I, Kohlbacher O, Kuepfer L, Lammert F, Maier D, Pfeifer N, Radde N, Rehm M, Roeder I, Saez-Rodriguez J, Sax U, Schmeck B, Schuppert A, Seilheimer B, Theis FJ, Vera J, and Wolkenhauer O (2018) Whither systems medicine? *Experimental & Molecular Medicine* 50(3), e453.
- Babaei S, Mahfouz A, Hulsman M, Lelieveldt BPF, de Ridder J, and Reinders M (2015) Hi-C chromatin interaction networks predict co-expression in the mouse cortex. *PLoS Computational Biology* 11(5): e1004221.
- Barabási A-L, Gulbahce N, and Loscalzo J (2011) Network medicine: A network-based approach to human disease. *Nature Reviews Genetics* 12(1): 56–68.
- Bergthaler A and Menche J (2017) The immune system as a social network. *Nature Immunology* 18(5): 481–482.
- Betsch C, Böhm R, Korn L, and Holtmann C (2017) On the benefits of explaining herd immunity in vaccine advocacy. *Nature Human Behaviour* 1(3): 1–6.
- Bianconi G and Barabási A-L (2001) Competition and multiscaling in evolving networks. *Europhysics Letters (EPL)* 54(4): 436–442.
- Bock DD, Lee W-CA, Kerlin AM, Andermann ML, Hood G, Wetzel AW, Yurgenson S, Soucy ER, Kim HS, and Reid RC (2011) Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471(7337): 177.
- Briggman KL, Helmstaedter M, and Denk W (2011) Wiring specificity in the direction-selectivity circuit of the retina. *Nature* 471(7337): 183.
- Brockmann D (2017) Public health: This message must be herd. *Nature Human Behaviour* 1(3): 0065.
- Brockmann D and Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342(6164): 1337–1342.
- Bullmore E and Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3): 186–198.
- Caldera M, Buphamalai P, Müller F, and Menche J (2017) Interactome-based approaches to human disease. *Current Opinion in Systems Biology* 3: 88–94.
- Caldera M, Müller F, Kaltenbrunner I, Licciardello MP, Lardeau C-H, Kubicek S, and Menche J (2019) Mapping the perturbome network of cellular perturbations. *Nature Communications* 10(1): 5140.
- Cheng F, Kovács IA, and Barabási A-L (2019) Network-based prediction of drug combinations. *Nature Communications* 10(1): 1197.
- Choudhary C and Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular Cell Biology* 11(6): 427–439.
- Christakis NA and Fowler JH (2007) The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357(4): 370–379.
- Christakis NA and Fowler JH (2008) The collective dynamics of smoking in a large social network. *New England Journal of Medicine* 358(21): 2249–2258.
- Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, Pelechano V, Styles EB, Billmann M, Van Leeuwen J, Van Dyk N, Lin ZY, Kuzmin E, Nelson J, Piotrowski JS, Srikumar T, Bahr S, Chen Y, Deshpande R, Kurat CF, Li SC, Li Z, Usaj MM, Okada H, Pascoe N, Luis BJS, Sharifpoor S, Shuteriqi E, Simpkins SW, Snider J, Suresh HG, Tan Y, Zhu H, Malod-Dognin N, Janjic V, Przulj N, Troyanskaya OG, Stagljar I, Xia T, Ohya Y, Gingras AC, Raught B, Boutros M, Steinmetz LM, Moore CL, Rosebrock AP, Caudy AA, Myers CL, Andrews B, and Boone C (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353(6306): aaf1420.
- Costanzo M, Kuzmin E, van Leeuwen J, Mair B, Moffat J, Boone C, and Andrews B (2019) Global genetic networks and the genotype-to-phenotype relationship. *Cell* 177(1): 85–100.
- Erdős P and Rényi A (1960) On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5(1): 17–60.
- Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, and D’Eustachio P (2016) The reactome pathway knowledgebase. *Nucleic Acids Research* 44(D1): D481–D487.
- Fischer B, Sandmann T, Horn T, Billmann M, Chaudhary V, Huber W, and Boutros M (2015) A map of directional genetic interactions in a metazoan cell. *eLife* 4.
- Fowler JH and Christakis NA (2009) Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ (Online)* 338(7685): 23–26.
- Ghiassian SD, Menche J, and Barabási A-L (2015) A Disease Module Detection (DIAMOND) Algorithm Derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Computational Biology* 11(4): e1004120.
- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615): 171–178.
- Greene JA and Loscalzo J (2017) Putting the patient back together—Social medicine, network medicine, and the limits of reductionism. *New England Journal of Medicine* 377(25): 2493–2499.
- Guney E, Menche J, Vidal M, and Barabási A-L (2016) Network-based in silico drug efficacy screening. *Nature Communications* 7(1): 10331.
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, and Neher RA (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34(23): 4121–4123.
- Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, and Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* 545(7655): 505–509.
- Jiang C, Wang X, Li X, Inlora J, Wang T, Liu Q, and Snyder M (2018) Dynamic human environmental exposome revealed by longitudinal personal monitoring. *Cell* 175(1): 277–291.
- Kalia V, Jones DP, and Miller GW (2019) Networks at the nexus of systems biology and the exposome. *Current Opinion in Toxicology* 16: 25–31.
- Karczewski KJ and Snyder MP (2018) Integrative omics for health and disease. *Nature Reviews Genetics* 19(5): 299–310.
- Kermack WO and McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 115(772): 700–721.
- Köhler S, Bauer S, Horn D, and Robinson PN (2008) Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics* 82(4): 949–958.

- Kovács IA, Luck K, Spirohn K, Wang Y, Pollis C, Schlabach S, Bian W, Kim D-K, Kishore N, Hao T, Calderwood MA, Vidal M, and Barabási A-L (2019) Network-based prediction of protein interactions. *Nature Communications* 10(1): 1–8.
- Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, Usaj M, Balint A, Usaj MM, Leeuwen Jv, Koch EN, Pons C, Dagilis AJ, Pryszyk M, Wang JZY, Hanchard J, Riggi M, Xu K, Heydari H, Luis B-JS, Shuteriqi E, Zhu H, Dyk NV, Sharifpoor S, Costanzo M, Loewith R, Caudy A, Bolnick D, Brown GW, Andrews BJ, Boone C, and Myers CL (2018) Systematic analysis of complex genetic interactions. *Science* 360(6386): eaao1729.
- Kveler K, Starosvetsky E, Ziv-Kenet A, Kalugny Y, Gorelik Y, Shalev-Malul G, Aizenbud-Reshef N, Dubovik T, Briller M, Campbell J, Rieckmann JC, Asbeh N, Rimar D, Meissner F, Wiser J, and Shen-Orr SS (2018) Immune-centric network of cytokines and cells in disease context identified by computational mining of PubMed. *Nature Biotechnology* 36(7): 651–659.
- Leskovec J and Sosič R (2016) SNAP: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8(1): 1.
- Lo Y-C and Torres JZ (2016) Chemical similarity networks for drug discovery. In: *Special Topics in Drug Discovery*. InTech.
- Longini IM, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DAT, and Halloran ME (2005) Containing pandemic influenza at the source. *Science (New York, N.Y.)* 309(5737): 1083–1087.
- Maslov S and Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296(5569): 910–913.
- McGillivray P, Clarke D, Meyerson W, Zhang J, Lee D, Gu M, Kumar S, Zhou H, and Gerstein MB (2018) Network analysis as a grand unifier in biomedical data science. *Annual Review of Biomedical Data Science* 1(1): 153–180.
- Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, and Barabási AL (2015) Uncovering disease-disease relationships through the incomplete interactome. *Science* 347(6224): 841.
- Meyer MJ, Beltrán JF, Liang S, Fragoza R, Rumack A, Liang J, Wei X, and Yu H (2018) Interactome INSIDER: A structural interactome browser for genomic studies. *Nature Methods* 15(2): 107–114.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, and Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298(5594): 824–827.
- Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatri-Aryamontri A, Dolinski K, and Tyers M (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Research* 47(D1): D529–D541.
- Pastor-Satorras R, Castellano C, Van Mieghem P, and Vespignani A (2015) Epidemic processes in complex networks. *Reviews of Modern Physics* 87(3): 925–979.
- Piñero A, Berenstein A, Gonzalez-Perez A, Chermomoretz A, and Furlong LJ (2016) Uncovering disease mechanisms through network biology in the era of Next Generation Sequencing. *Scientific Reports* 6(1): 24570.
- Piotrowski JS, Li SC, Deshpande R, Simpkins SW, Nelson J, Yashiroda Y, Barber JM, Safizadeh H, Wilson E, Okada H, Gebre AA, Kubo K, Torres NP, LeBlanc MA, Andrusiak K, Okamoto R, Yoshimura M, DeRango-Adem E, van Leeuwen J, Shirahige K, Baryshnikova A, Brown GW, Hirano H, Costanzo M, Andrews B, Ohya Y, Osada H, Yoshida M, Myers CL, and Boone C (2017) Functional annotation of chemical libraries across diverse biological processes. *Nature Chemical Biology* 13(9): 982–993.
- Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, Ono K, Miello C, Hicks L, Szalma S, et al. (2015) Ndx, the network data exchange. *Cell Systems* 1(4): 302–305.
- Rauscher B, Heigwer F, Henkel L, Hielscher T, Voloshanenko O, and Boutros M (2018) Toward an integrated map of genetic interactions in cancer cells. *Molecular Systems Biology* 14(2): e7656.
- Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrossay D, Sallusto F, Shen-Orr SS, Lanzavecchia A, Mann M, and Meissner F (2017) Social network architecture of human immune cells unveiled by quantitative proteomics. *Nature Immunology* 18(5): 583–593.
- Rolland T, Taşan M, Charleatoux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kirov R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabási AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, and Vidal M (2014) A proteome-scale map of the human interactome network. *Cell* 159(5): 1212–1226.
- Saha A, Kim Y, Gewirtz AD, Jo B, Gao C, McDowell IC, Engelhardt BE, Battle A, and Battle A (2017) Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Research* 27(11): 1843–1858.
- Sporns O (2013) The human connectome: origins and challenges. *NeuroImage* 80: 53–61.
- Stéphanou A, Fanchon E, Innominato PF, and Ballesta A (2018) Systems biology, systems medicine, systems pharmacology: The what and the why. *Acta Biotheoretica* 1–21.
- Veneman WJ, Spaink HP, Brun NR, Bosker T, and Vijver MG (2017) Pathway analysis of systemic transcriptome responses to injected polystyrene particles in zebrafish larvae. *Aquatic Toxicology* 190: 112–120.
- Wang Z, Moreno Y, Boccaletti S, and Perc M (2017) Vaccination and epidemics in networked populations—An introduction. *Chaos, Solitons and Fractals* 103: 177–183.
- White JG, Southgate E, Thomson JN, and Brenner S (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 314(1165): 1–340.
- Zitnik M, Sosič R, Maheshwari S, and Leskovec J (2018) BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. <http://snap.stanford.edu/biodata>.

Relevant Websites

Websites for delving deeper into network theory, including many visualizations and applications from different areas and interactive examples

<http://networksciencebook.com/>.

<http://www.complexity-explorables.org/>.

Useful resources to build or annotate biological networks

<http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/>.

<https://thebiogrid.org/>.

<http://www.disgenet.org/>.

<https://gtexportal.org>.

<http://snap.stanford.edu/data/>.

1.3.3 Network medicine as a framework

From its inception, the field of network medicine has aimed at elucidating “*the molecular relationships between apparently distinct (patho)phenotypes*”, as defined by Barabási and colleagues (Barabási et al., 2011). This can indeed be achieved as the genes mutated in inherited diseases display specific properties in molecular interaction networks (Feldman et al., 2008; Goh et al., 2007). In particular, the well-studied localization of disease-related genes in network modules highlights the relevance of the approach to explore the molecular mechanisms underlying pathologies (Ghiassian et al., 2015; Menche et al., 2015). Genes implicated in drug response, adverse effect and modulating pharmacokinetics parameters are also associated with various multi-scale network structures (Piñero et al., 2016). Such network-informed principles were further leveraged for biomedical applications. Notably, biological interaction networks proved to be particularly adapted to support drug repurposing, an approach in which drugs that are already in clinical use for a given disease are investigated for the treatment of another disease (Cheng et al., 2018; Guney et al., 2016). If successful, drug repurposing would drastically decrease drug development and treatment costs by avoiding compound optimization and safety trials, and using existing production sites. The approach has been used among others for neurological disorders (Lüscher Dias et al., 2020) and COVID-19 (Morselli Gysi et al., 2021; Verstraete et al., 2020; Zhou et al., 2020). On top of the economical benefits of drug repurposing, the outbreak of infectious diseases such as COVID-19 requires a rapid response, and considerable time could be gained by reusing approved drugs with established supply chains (Sharma et al., 2021). Because networks excel at representing how novel properties emerge from the interaction between individual elements, network medicine also delivered new results on the modelling and prediction of the effect of combinations between multiple drugs (Caldera et al., 2019; Cheng et al., 2019; Hu et al., 2019; Regan-Fendt et al., 2019). Accordingly, the advances of polypharmacology, defined as the synergistic use of multiple treatments, increasingly allow to tackle causal disease mechanisms instead of treating symptoms (Nogales et al., 2021). Biological networks can also be used in multiple other ways, for instance to contextualize the transcriptomic profiles of cardiomyopathy patients (Maron et al., 2021), to better translate knowledge gained in animal models to human physiology (Blais et al., 2017), or for compound prioritization in drug discovery (Sidders et al., 2018). As cancer is among the leading causes of death in upper income countries and is primarily a somatic genetic disease, it is naturally one of the diseases studied the most at the molecular level. Thus, multiple studies aimed to leverage network approaches to offer more personalized treatments (Dinstag & Shamir, 2020; Liu et al., 2020; Zhang et al., 2017). More recently, protein interaction have been mapped in detail and compared in the context of different cancer types. Specific protein interaction interfaces were found to be disrupted by common cancer mutations, and this finding provided suggestions for potential novel therapies (Kim et al., 2021; Swaney et al., 2021; Zheng et al., 2021).

These last examples highlight the plasticity of protein-protein interactions (PPIs), which is an important observation for future research directions. Indeed, most studies used general-purpose

interactomes derived for the whole genome of model cell lines using yeast two-hybrid or mass spectrometry methods so far (Luck et al., 2020; Richards et al., 2021). An alternative is to take a patient's or a disease's mutation and expression profiles into account to infer a customized PPI network that better represents the interactions happening in a context of interest. Such networks are typically assembled under the assumption that proteins would interact if they are known to do so in model cell lines and if both interaction partners are abundant in a particular context, which is often proxied by the expression level of the corresponding genes (Liu et al., 2014). Another layer of complexity arises from the tissue specificity of certain interactions. Considerable efforts have been made to map transcriptome levels and splicing across biological tissues, most recently by the GTEx consortium (Melé et al., 2015). Tissue-specific networks revealed that PPIs differ to a large extent between tissues, with a trend for tissue-specific genes to interact with essential housekeeping genes (Bossi & Lehner, 2009; Yeger-Lotem & Sharan, 2015). These findings led to improvements in disease gene prioritization (Magger et al., 2012). Expression levels can also be considered to define networks of disease-tissue associations, based on the local presence of disease proteins and their interactors (Kitsak et al., 2016). Taken together, these studies reinforce the need to further depart from model cell lines, and study which interactions are conserved or experiment-specific. Then, one can generalize and validate such findings, so that network-based methods can be more adapted to various biological contexts. Another evolution of network medicine might come from the increased tendency of biological data to be thought of and natively stored as entities connected in a network through different types of relationships. This emerging concept is referred to as knowledge graphs and is not limited to biomedical applications (Hogan et al., 2022). Several projects already compiled biological and clinical information into large knowledge graphs that can be queried to answer complex questions (Santos et al., 2020; Walsh et al., 2020). For instance, they were used to ease the identifications of pathways, drugs and biomarkers. Another noticeable example comes from the gene ontology consortium, which is already well established for their extensive classification of cellular processes and their functional annotation of genes (Ashburner et al., 2000). They recently implemented a new framework called Causal Activity Modeling, aiming to be more flexible and to better structure the relationships between different gene annotations compared, which can be seen as a knowledge graph (Thomas et al., 2019).

The development of network medicine is also tightly linked to the emergence of modern molecular biology techniques, in particular high-throughput sequencing and its applications (Goodwin et al., 2016). However, the considerable advances provided by the so-called OMICS technologies in biomedical research were not followed by a large adoption for clinical translation so far (Karczewski & Snyder, 2018). A major open challenge is to better integrate the complementary aspects provided by multiple OMICS modalities to form a more comprehensive description of a disease phenotype and deliver actionable patient-specific insight. From their most basic definition, networks model connections between agents, without enforcing what the connections or agents represent. This means that they inherently offer a solution to compile different types of relationships together if deemed meaningful, and therefore provide a way to integrate data

generated in multiple OMICS experiments (Tuncbag et al., 2016). Different biological processes can be seen and modelled as a network of heterogeneous networks, each one describing a subsystem such as a tissue, or using a different definition for edges. Theoretical frameworks also exist to analyse such a network of networks (Gao et al., 2014; Kawakubo et al., 2019). In practice, various molecular information, ranging from co-expression and shared drug association to physical interactions and pathway membership, can also be combined into large multilayer networks. This approach can reveal disease signatures and can be used to progressively refine which type of information is most relevant to characterize a particular disease (Buphamalai et al., 2021). Linking such disease-related gene networks also allows for patient stratification and for the identification of groups of genes with a functional disease relevance (Núñez-Carpintero et al., 2021).

Finally, not only traditional OMICS modalities such as transcriptomics, metabolomics and proteomics served to build informative biological networks. At multiple occasions, networks describing how genetic or chemical interactions combine were assembled based on HCS experiments with an imaging readout (Caldera et al., 2019; Fischer et al., 2015; Heigwer et al., 2018; Mattiazzi Usaj et al., 2020). One key overarching question addressed in the work compiled in this thesis is precisely how to best extract morphological profiles from such experiments so that they can characterize disease phenotype and be integrated with prior information through networks.

Aims of the thesis

The work described in this thesis aims first and foremost at developing and applying new ways to compile, curate, contextualize and interpret morphological profiles in high-content imaging (HCI) experiments. Indeed, new ways of analysing HCI data would serve as a computational and analytical basis for the systematic characterization of perturbations of biological systems. In particular, we seek to be able to integrate morphological profiles with other data types, following the guiding principles of systems medicine. Figure 2.1 shows how our contributions to morphological profiling and its integration with systems medicine approaches could increase the biological insight that can be gained from perturbation-based biological experiments. Our working hypothesis is twofold. First, improved methods for morphological profiling would allow to extract more information from the imaging readout than the more targeted analyses that are currently used the most. Second, a better integration of morphological data within the framework of network medicine could add up to the knowledge derived from other experimental modalities, with concrete biomedical and clinical applications.

This thesis started with an introduction to the ways in which biological systems can be probed via multiple types of perturbations and how more complex and powerful screening setups were progressively developed to evaluate the effects of such perturbations. The diversity of these setups is also linked to the flexibility that screens enable in terms of scale, readout and types of perturbations. The emergence and current state of HCI screens and associated analysis methods is then summarized. I then highlighted how these technologies enable large-scale screening experiments with multidimensional and functional readouts, which are ideal for systems biology studies. By insisting on the potential of HCS experiments, we motivate our choice of conducting large-scale imaging screens to explore guiding principles of cell biology. Our decision is also substantiated by a more precise introduction of the core computational concepts relevant to the morphological profiling analyses reproduced later in this document, such as the data transformation and distance computation steps. I also provided an overview

of network medicine techniques, by including a literature review, authored by my doctoral supervisor and me, and published as a book chapter. This review is complemented by a broader description of the systems biology paradigm and of the current achievements and promises of network medicine. Such network-based techniques are suitable to many experimental modalities and are also highly relevant to the topic of HCI screens, as the resulting morphological profiles are best contextualized in the light of the prior knowledge available about the biological systems studied. The work described in this thesis suggests that network-based integration methods are especially useful to achieve this goal. We have now set up the context and stake of image-based profiling analyses. In the following chapter, I will present concrete advances in the direction of the aforementioned goals, with both methodological improvements and novel biological results, in the form of two research articles.

The first one is entitled “BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations”. Presented in section 3.1, it showcases a software solution for handling profiles that compile informative descriptors of cellular morphology in different experimental conditions. It also introduces a way to compare these profiles via robust statistical distances, and demonstrates that they can further be integrated with external perturbation annotations. In the analysis conducted, drug profiles are put into context with a comparison to a biological network representing protein-protein interactions.

The second article, entitled “Morphological profiling of human T and NK lymphocytes by high-content cell imaging”, is included in section 3.2. It relies on the HCI technology and on morphological profiles to study the structure of the immunological synapse in cytotoxic lymphocytes. The main experiments involved the assembly and comparison of the morphological profiles for different disruptions of the normal actin organization. With this, we observed the structure of the actin cytoskeleton and the distribution of lytic granules in three dimensions, in healthy cells and in the context of inborn errors of immunity. Notably, we took different approaches in our two compiled studies to interrogate morphological profiles and to determine what morphological properties are linked to specific experimental conditions. In the first one, we looked at the morphological measurements changing the most in different conditions. In contrast, we used machine learning in the second study to attribute morphological properties to actin defects. Moreover, this article also illustrates that HCI can serve as a tool for hypothesis generation, guiding further complementary experiments, such as live cell imaging and degranulation assays. Finally, this thesis includes an extensive discussion of the implications and limitations of our results and the avenues they opened. This includes both observations that were made when working on the two compiled research articles, as well as open challenges in the fields of morphological profiling and network medicine that we believe could be further addressed. Accordingly, current work is ongoing in our research group which builds upon our previous achievements to study via HCI the effect of large libraries of genetic and chemical perturbations, in isolation and in combination. This discussion closes with an assessment of the progress we and others made regarding the analysis and contextualization of morphological profiling. In particular, we see how high-content imaging, morphological profiling and network medicine fit together, and

we figure out where we stand regarding clinical translation.

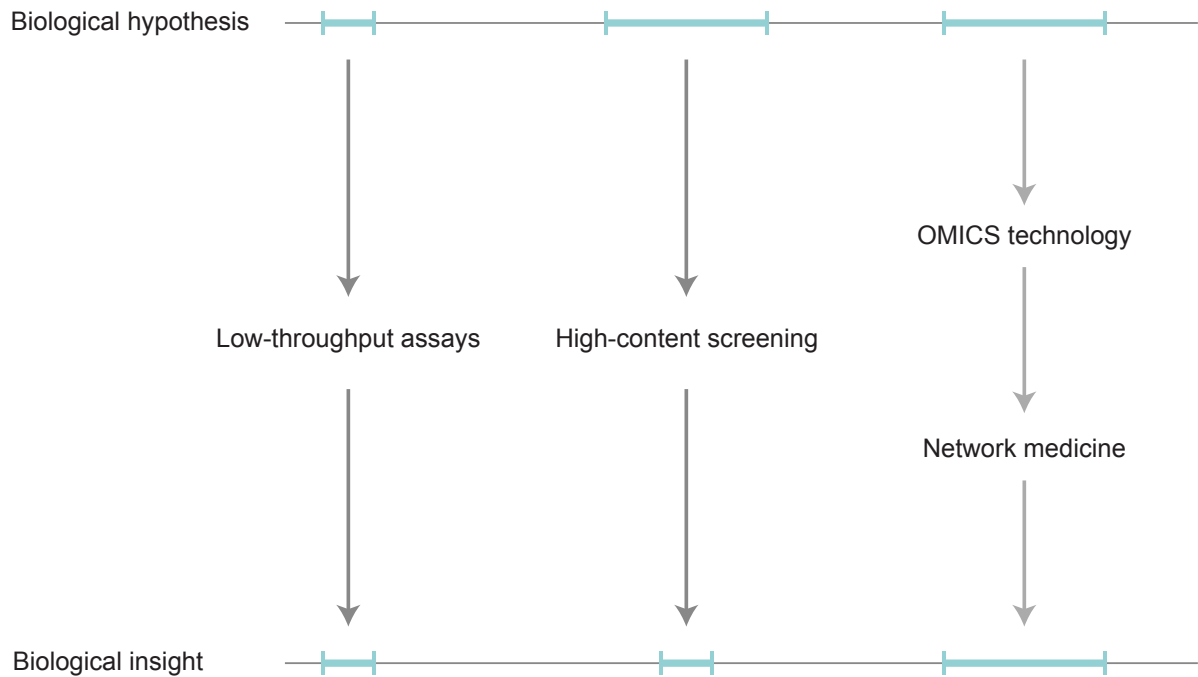
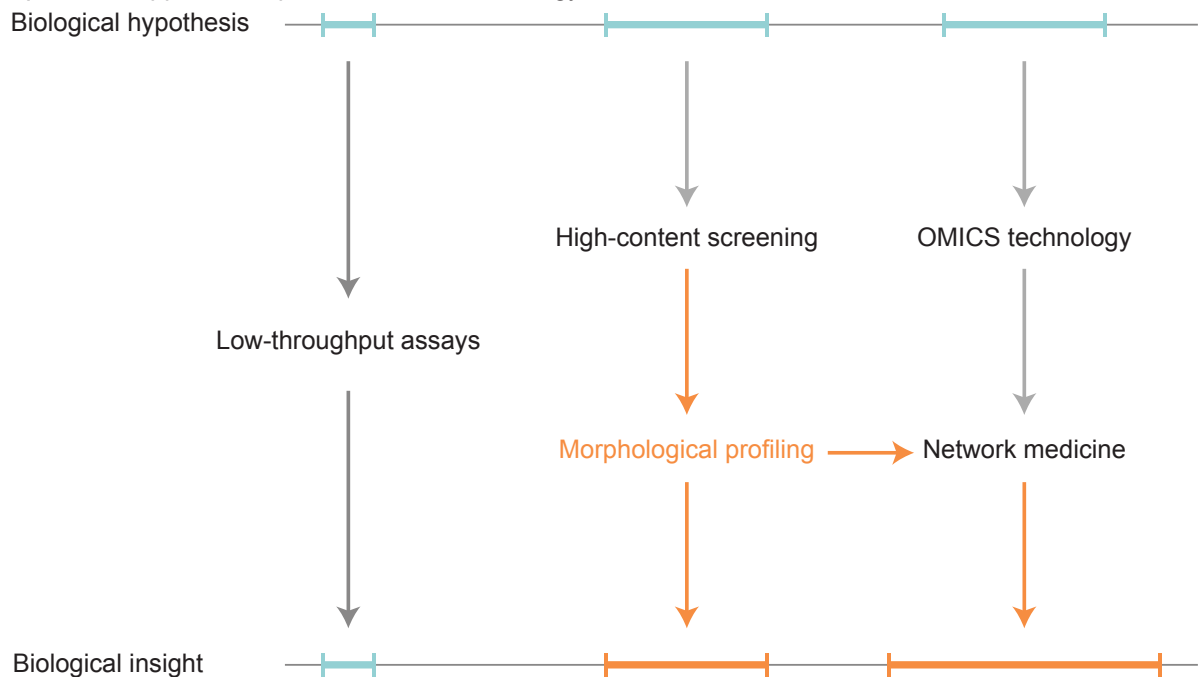
a) Standard approach to perturbation-based biology**b) Modern approach to perturbation-based biology**

Figure 2.1: Potential contribution of morphological profiling and network medicine to research in molecular biology. In **a)** the traditional approach to biological perturbation experiments, low-throughput assays focus on specific targets and only provide information about a restricted number of molecules or mechanisms. For one thing, high-content assays allow to test more perturbations simultaneously, but are often underused, for instance by only serving as a proxy for cell viability, therefore providing limited insight. At the same time, OMICS technologies can also be high-throughput and provide a rich readout, which can be leveraged by network medicine to gain system-level insight. This thesis aims to contribute to the emergence of **b)** a new paradigm for HCS analysis, in which informative morphological profiles are systematically assembled. They would unlock the full potential of the data generated in HCS experiments, by being able to represent quantitatively all measurable phenotypic changes. These image-based profiles would also constitute an additional layer of information in systems medicine approaches. Thus, they would complement well-established OMICS technologies, and increase even more the insight that network medicine may provide and its potential for clinical translation.

3.1 BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations

Loan Vulliard, Joel Hancock, Anton Kamnev, Christopher W. Fell, Joana Ferreira da Silva, Joanna I. Loizou, Vanja Nagy, Loïc Dupré and Jörg Menche.

Published in *Bioinformatics*. Volume 38, issue 8, pages 1692-1699. Online publication date: December 22, 2021. Print publication date: March 15, 2022.

DOI: <https://doi.org/10.1093/bioinformatics/btab853>.

As introduced in section 1.2.5, morphological profiling consists in creating and comparing profiles describing cellular morphology assessed in imaging screens. This approach has the potential to systematically characterize a wide range of biological and experimental conditions in multiple organisms and cell types (Breinig et al., 2015; Caldera et al., 2019; Fischer et al., 2015; Heigwer et al., 2018). However, it requires the careful selection and transformation of the initial quantitative descriptors extracted from the images. Well-established guidelines enumerate which tasks typically need to be performed (Bougen-Zhukov et al., 2017; Caicedo et al., 2017) but there is currently no flexible standard implementation adopted by the research community. In particular, no solution was available for the up-and-coming programming language Julia (Bezanson et al., 2017; Roesch et al., 2021). We aimed to fill this gap with BioProfiling.jl, a toolkit covering different aspects of morphological profiling, ranging from data normalization and filtering to visualization and quantification of the statistical significance of morphological changes. In the following paper, we present what BioProfiling.jl can achieve.

To do so, we analyze a compound screen and compare the information gained about drug mechanisms of action either from the morphological changes they induce or from the network of interactions between the proteins they target.

When working on multiple morphological analyses run in our group, we noticed the lack of flexible and efficient tools to ease the handling and filtering of morphological profiles. This prompted the development of BioProfiling.jl. The package started as a compilation of methods that I implemented, in part for the analyses I produced for the study presented in section 3.2, and sought to made reusable by myself and others. It was then expanded and improved with additional methods to make it an end-to-end solution able to handle “raw” morphological measurements and transform them into profiles usable for multiple tasks. This is achieved with a unified conceptual framework, in which the data is wrapped in a dedicated object and progressively curated. Keeping a high degree of control over the filtering and data transformation process allowed us to explore multiple strategies to define what constitutes a significant morphological activity. We observed that the choice of profiling method greatly affect the final results. In our main approach, we used robust statistics to account for the presence of imaging artifacts and biological outliers and developed a robust distance inspired by the Hellinger distance which resulted in a meaningful list of hit compounds.

All the code of the package was written by myself, and is shared as an open source project on the code sharing platform GitHub. The features, concepts and directions given to the project were discussed together with Joel Hancock and Prof. Jörg Menche. All other authors were involved in the design, supervision and realization of the drug screen. Of note, we are closely collaborating with the labs led by Dr. Vanja Nagy, Dr Loïc Dupré and Ass.-Prof. Joanna Loizou for the development of large image-based screens, described in the Discussion section of this thesis. The main dataset used in the article below was generated in order to explore the morphological effects of chemical compounds in the context of a larger chemical-genetic screen. The microscopy images as well as the phenotypic measurements were made freely available for other researchers to use, both for method development and to further explore the morphological activity of the compounds. Finally, the resulting manuscript was written by Prof. Jörg Menche and myself, with contributions from all co-authors.

Supplementary material can be found in the appendix chapter at the end of this thesis.

Bioimage informatics

BioProfiling.jl: profiling biological perturbations with high-content imaging in single cells and heterogeneous populations

Loan Vulliard^{1,2}, Joel Hancock^{1,2}, Anton Kamnev^{3,4}, Christopher W. Fell^{1,3,5}, Joana Ferreira da Silva^{1,6}, Joanna I. Loizou^{1,6}, Vanja Nagy^{1,3,5}, Loïc Dupré^{3,4,7} and Jörg Menche^{1,2,8,*} 

¹CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna 1090, Austria, ²Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Vienna 1030, Austria, ³Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna 1090, Austria, ⁴Department of Dermatology, Medical University of Vienna, Vienna 1090, Austria, ⁵Department of Neurology, Medical University of Vienna, Vienna 1090, Austria, ⁶Institute of Cancer Research, Department of Medicine I, Medical University of Vienna and Comprehensive Cancer Center, Vienna 1090, Austria, ⁷Toulouse Institute for Infectious and Inflammatory Diseases (INFINITY), INSERM UMR1291, CNRS UMR5051, Toulouse III Paul Sabatier University, Toulouse 31024, France and ⁸Faculty of Mathematics, University of Vienna, Vienna 1090, Austria

*To whom correspondence should be addressed.

Associate Editor: Olga Vitek

Received on June 15, 2021; revised on December 15, 2021; editorial decision on December 15, 2021; accepted on December 16, 2021

Abstract

Motivation: High-content imaging screens provide a cost-effective and scalable way to assess cell states across diverse experimental conditions. The analysis of the acquired microscopy images involves assembling and curating raw cellular measurements into morphological profiles suitable for testing biological hypotheses. Despite being a critical step, general-purpose and adaptable tools for morphological profiling are lacking and no solution is available for the high-performance Julia programming language.

Results: Here, we introduce BioProfiling.jl, an efficient end-to-end solution for compiling and filtering informative morphological profiles in Julia. The package contains all the necessary data structures to curate morphological measurements and helper functions to transform, normalize and visualize profiles. Robust statistical distances and permutation tests enable quantification of the significance of the observed changes despite the high fraction of outliers inherent to high-content screens. This package also simplifies visual artifact diagnostics, thus streamlining a bottleneck of morphological analyses. We showcase the features of the package by analyzing a chemical imaging screen, in which the morphological profiles prove to be informative about the compounds' mechanisms of action and can be conveniently integrated with the network localization of molecular targets.

Availability and implementation: The Julia package is available on GitHub: <https://github.com/menche/bioprofiling>. We also provide Jupyter notebooks reproducing our analyses: <https://github.com/menche/bioprofilingnotebooks>. The data underlying this article are available from FigShare, at <https://doi.org/10.6084/m9.figshare.14784678.v2>.

Contact: joerg.menche@univie.ac.at

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

High-Content Screening (HCS) enables profiling cellular phenotypes across hundreds of thousands of conditions by combining

automated microscopy with advanced image analysis methods. HCS thus represents a flexible and cost-effective solution for replacing multiple specific assays (Chandrasekaran *et al.*, 2020; Simm *et al.*, 2018; Way *et al.*, 2021a), and has been widely adopted in both basic

and applied research. Notable achievements range from drug discovery (Chandrasekaran et al., 2020; Scheeder et al., 2018; Simm et al., 2018) to the elucidation of combinatorial drug effects (Caldera et al., 2019) and to *ex-vivo* drug-response screening in patients (Snijder et al., 2017). Depending on the application, the analysis of HCS experiments may involve a variety of tasks. For instance, one might perform a classification task to infer the mechanism of action of candidate drugs (Ando et al., 2017; Ljosa et al., 2013; Pawlowski et al., 2016), compare cellular phenotypes in various conditions (German et al., 2021; Gustafsdottir et al., 2013; Rohban et al., 2017) or describe interactions between cellular perturbations (Billmann et al., 2016; Breinig et al., 2015; Caldera et al., 2019; Fischer et al., 2015; Heigwer et al., 2018). All these cases involve numerous experimental and analytical steps.

A typical HCS experiment starts from preparing microplates with cells subjected to various perturbations, such as different drugs, and stained using standardized protocols, such as the Cell Painting assay (Bray et al., 2016; Fig. 1a). These microplates are then imaged using automated confocal fluorescence microscopy, resulting in a large number of images. Each image is then analyzed to extract quantitative morphological measurements that describe the respective cellular phenotype. Some tools are commonly used for this numerical feature extraction step (McQuin et al., 2018; Pau et al., 2010), and recent deep learning approaches attempt to replace expert-curated measurements with data-driven discriminative features (Ando et al., 2017; Lu et al., 2019; Pawlowski et al., 2016).

While the analytical tasks of HCS experiments vary between applications, they involve common data normalization and filtering steps, and guidelines have been proposed for computing informative representations of cellular phenotypes, usually referred to as morphological profiles (Bougen-Zhukov et al., 2017; Caicedo et al., 2017). An analysis pipeline suitable to facilitate morphological profiling should meet several criteria. First, it should be versatile, to adapt to different HCS use cases and to cope with the diverse challenges inherent to such experiments (Boutros et al., 2015; Caicedo et al., 2017; Chandrasekaran et al., 2020; Ljosa et al., 2013). These challenges include technical problems such as blurred images, poorly adherent cells, saturated pixels, staining artifacts and segmentation mistakes. HCS studies need to address these frequent limitations, as in some experiments most images are affected (Fig. 1b and c). Second, the approach should account for background noise, intensity bias and potential confounders, including plate layout and batch effects. Third, the considerable heterogeneity of the morphological descriptors needs to be handled. Cellular morphology might vary greatly in the analyzed cell populations due to the experimental setup, heterogeneous cell types or cell states, inconsistencies in perturbation efficiency, or when timing-dependent phenomena are imaged as snapshots.

The few actively maintained HCS analysis tools attempting to fulfill these needs include CellProfiler Analyst and its graphical user interface, designed to handle CellProfiler measurements (Jones et al., 2008; McQuin et al., 2018), as well as the general-purpose cytominer and Pycytominer in the R and Python languages, respectively (Becker et al., 2021; Way et al., 2021b). There are also packages addressing similar challenges but focusing on other modalities (which generally provide less spatial information or less throughput than high-content imaging screens) such as the R packages cellHTS2, optimized for measurements from plate readers (Boutros et al., 2006), and more recently cytomapper for analyzing imaging mass cytometry experiments (Eling et al., 2021). Despite the existence of these tools, HCS analysts still heavily rely on custom implementations of morphological profile curation for each study to account for different imaging modalities and analytical goals (Ziegler et al., 2021).

Julia is a high-performance, high-level open-source programming language specifically designed for scientific computing and data science (Bezanson et al., 2017). It is increasingly adopted by researchers in bioinformatics and biomedical research (Roesch et al., 2021), with applications ranging from protein sequence analysis (Zea et al., 2016) to structural bioinformatics (Greener et al., 2020) and flux balance analysis (Heirendt et al., 2017). Julia is also ideal for tackling the

challenges of morphological analyses, as they are both computationally demanding and inherently high-level. In this article, we introduce BioProfiling.jl, the first Julia library for efficient and convenient morphological profiling that (i) handles noisy data through systematic filtering and robust statistics, (ii) provides dedicated functions to normalize data and mitigate layout effects and (iii) implements statistical tests for quantifying the strength of morphological changes that take the variability of morphological profiles into account. Our integrated software solution is thus bridging the existing gap between experimental data and biological interpretation. Furthermore, we conduct an image-based chemical screen to validate our approach and characterize the morphological impact of compounds in U-2-OS cells.

2 Materials and methods

2.1 Package implementation and features

We created BioProfiling.jl, a package for the Julia programming language that compiles over 30 methods and data structures for all steps in assembling and curating morphological profiles. To enable the bioimage analysis community to apply BioProfiling.jl to their own data, a complete documentation and a set of notebooks reproducing the analyses described in this paper are provided. In brief, the whole process of morphological profiling is conceptually simplified by defining an *Experiment* object that includes both quantitative data and metadata in a tabular format, and methods able to interact with these objects directly to curate, transform and visualize the corresponding profiles. After creating the *Experiment* object from a table of morphological features, such as measurements obtained from CellProfiler (McQuin et al., 2018) or activation values from a deep neural network (Ando et al., 2017; Lu et al., 2019; Pawlowski et al., 2016), one would typically filter entries (rows representing biological units) and select features (columns representing phenotypic descriptors) with the *Filter* and *Selector* types, respectively. Convenient shorthand is provided such as the *NameSelector* type to select features based on their name rather than their values, or the *CombinationFilter* type to join simple *Filter* objects with any logical operator. The selected measurements can then be transformed with the *logtransform!* and *normtransform!* methods, and *decorrelate!* discards highly correlated measurements. The filtered *Experiment* objects also support uniform manifold approximation and projection (UMAP) visualizations (McInnes et al., 2018) as implemented in UMAP.jl. The resulting feature profiles can be visually inspected by highlighting images and individual cells matching a *Filter* with the *diagnostic_images* method, currently implemented for TIFF images in any accessible folder. Up to three distinct files can be specified to produce an RGB image. Finally, *robust_morphological_perturbation_value* and efficient implementations of statistical distances, described in detail below, are available for quantifying the significance of morphological changes induced by a particular perturbation. Freely available from GitHub and the Julia package registry under the MIT license, BioProfiling.jl is part of a growing open-source software ecosystem ensuring that it stays flexible, maintainable and interoperable. To ensure its stability, the package is thoroughly validated with more than 120 tests, systematically run on multiple environments using GitHub Actions for continuous integration. The total testing coverage is reported using Codecov. Together with the simplicity of its design and properties of the Julia language itself such as multiple dispatch, BioProfiling.jl can easily be extended by users to address their specific use cases if they are not yet covered by the features we implemented. Finally, we encourage such contribution to be integrated and shared through pull requests on the BioProfiling.jl repository.

2.2 Cell culture

We selected the U-2-OS cell line as it is morphologically expressive and commonly used in HCS experiments (Gustafsdottir et al., 2013; Rohban et al., 2017; Wawer et al., 2014). U-2-OS cells (ATCC HTB-96) were cultured in high glucose Dulbecco's modified Eagle's medium (Thermo Fisher #11960044), 10% fetal bovine serum (Sigma-Aldrich #F0804), 1× penicillin/streptomycin

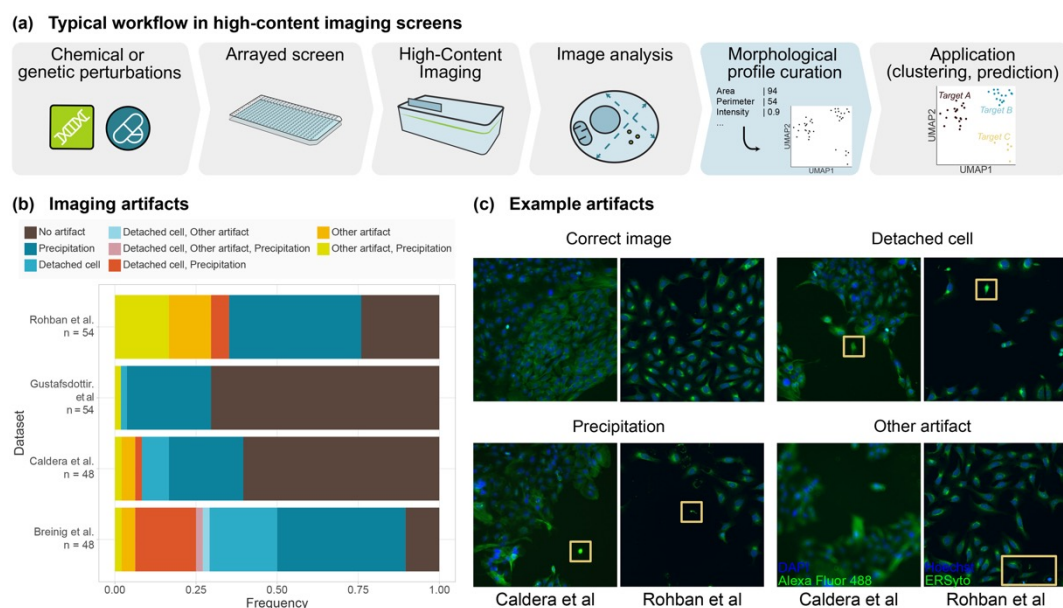


Fig. 1. HCS experiments require adequate analysis tools. (a) Standard analysis workflow of HCS experiments. (b) Quantification of imaging artifacts that may lead to biases in HCS analyses in sample images from four published studies (Breinig et al., 2015; Caldera et al., 2019; Gustafsdottir et al., 2013; Rohban et al., 2017). (c) Examples of such imaging artifacts. Boxes highlight regions of interest

(Biowest #L0022-020) and 1 mM sodium pyruvate (Thermo Fisher #11360070) and maintained in a humidified incubator (5% CO₂, 37°C).

2.3 Chemical screen

A total of 311 compounds were selected to cover a wide range of biological processes and based on their propensity to impact cellular morphology in in-house and published studies (Wawer et al., 2014). A full list of the compounds and their concentration is provided in [Supplementary Table S1](#). Drugs were transferred to 384-well plates (PerkinElmer #6057302) using a liquid handler, in which 32 dimethyl sulfoxide (DMSO) wells were used as a reference to assess the effect of the compounds as DMSO was used as solvent for the chemical library. The positions of the compounds were randomized on each plate while ensuring the presence of two DMSO control wells in each row and of one to two in each column. Two drug plates were seeded in 50 µl of culture medium with U-2-OS cells at 750 and 1500 cells/well, respectively, and incubated at 37°C with 5% CO₂ for 72 h. Living cells were then washed three times with phosphate-buffered saline (PBS) and stained for 10 min using CellMask Orange Plasma membrane stain (Thermo Fisher #C10045). Cells were washed three more times with PBS and fixed with a solution of 4% Formaldehyde (Thermo Fisher #28908). After washing three more times with PBS, cells were permeabilized with 50 µl of permeabilization solution, consisting of PBS supplemented with 0.1% saponin-based permeabilization solution (Invitrogen #00-8333-56) and 5% fetal calf serum (Sigma #F7524), for 1 h. F-actin was stained overnight with Phalloidin-488 staining solution (0.6 U/ml in permeabilization buffer; Thermo Fisher #A12379). Nucleic acids were stained with 30 µl of 4',6-diamidino-2-phenylindole (DAPI, 5 µg/ml in PBS, Thermo Fisher #D1306) for 10–20 min. Finally, cells were washed three times with PBS and 50 µl of PBS solution was added per well. The entire surface of each well was imaged (20 fields of view with a 20× magnification Long-Working Distance (LWD) objective) on an Operetta High-Content Imaging System (PerkinElmer) using three fluorescence channels to detect DAPI (360–400/410–480 nm), Phalloidin (460–490/500–550 nm) and CellMask (520–550/560–630 nm). All images are available from FigShare (DOI: 10.1101/2021.06.18.448961).

2.4 Image analysis

We processed and analyzed microscopy images using CellProfiler 3.1.8 (McQuinn et al., 2018), the full pipeline is available from FigShare (DOI: 10.1101/2021.06.18.448961). In brief, the image quality was assessed, the intensities were log-transformed, the illumination on each image was corrected based on background intensities before segmenting cell nuclei using global minimum cross entropy thresholding. Two successive secondary segmentation steps were performed using the propagation method (Jones et al., 2005) and global minimum cross entropy thresholding first on the CellMask then on the phalloidin channel to detect the cell bodies surrounding each nucleus. Finally, measurements were acquired for intensities in the nuclei and cytoplasms, granularity on all channels, textural and shape features, intensity distributions and number of neighboring cells <5 pixels away. This led to a total of 385 morphological features per cell.

2.5 Morphological profiling with BioProfiling.jl

All measurements were compiled for each cell in a BioProfiling.jl *Experiment* object, and non-numerical and uninformative features such as cell orientation were excluded from the profiles. We designed four cell filters to exclude technical outliers such as poorly segmented objects. Thresholds for the different filters were set by based on the distribution of geometrical and intensity measurements, so that extreme values would be discarded, while ensuring that the filtered objects were indeed problematic via a systematic visual inspection with the diagnostic tools available in BioProfiling.jl. These filters excluded cells with high CellMask to Phalloidin or DAPI to CellMask segmented area ratios, with a low nucleus form factor or with a high maximal CellMask intensity. From these curated measurements, we aggregated profiles for each field of view containing three valid cells or more by taking the median value of all cell-level values corresponding to each field of view, for each feature individually. We then removed features which were constant across all DMSO controls or over the complete plate, and log-transformed the values to reduce the skewness of the distribution of some measurements. To correct for plate effects and bring features to comparable scales, for each feature, individual field-of-view measurements were centered and scaled based on the median and median absolute deviation (MAD) of the control profiles in the same row or column, as follows:

$$\hat{s}_{x,y} = \frac{s_{x,y} - \text{median}(S_{ctrl_{xy}})}{\text{mad}(S_{ctrl_{xy}})},$$

with $s_{x,y}$ being the initial value of a field of view in row x and column y for feature s , $\hat{s}_{x,y}$ its value after correction, and $S_{ctrl_{xy}}$ the set of all values in control wells either in row x or column y . After this transformation, a high feature spread corresponds to a large deviation from the negative control profiles for some perturbations. We then reduced redundancy in the profiles by ordering features by decreasing MAD, which prioritizes features displaying changes compared to controls, and sequentially removing features with a Pearson's correlation coefficient higher than 0.8 with any of the previously selected features. We obtained the list of most variable selected features using the *most_variable_features* method (Supplementary Table S2). Lastly, we reduced the profiles to four dimensions with UMAP (McInnes et al., 2018), aiming to preserve the cosine distances between points, with *min_dist* set to 2 and all other parameters left to default values. We also visualized the results when no filters were applied using a two-dimensional UMAP embedding with default parameters and projecting cell-level profiles for all objects. We found several clusters driven by artifacts and biological outliers (Supplementary Fig. S1).

2.6 Hit detection with the robust Hellinger distance

BioProfiling.jl offers several statistical distances for quantifying the significance of morphological changes in HCS. The Mahalanobis distance takes the spread of the data in each dimension into account, which can be useful to compare two experimental conditions as previously described (Hutz et al., 2013). We also implemented the robust Mahalanobis distance which does not get biased by outliers by replacing the mean and covariance matrix by robust estimators of location and dispersion obtained using the minimum covariance determinant (MCD) algorithm (Cabana et al., 2019; Rousseeuw and van Driessen, 1999). This approach was used previously in an HCS analysis (German et al., 2021), yet without efficient, ready-to-use implementation. In comparison, the profile curation could be twice as compact when taking advantage of BioProfiling.jl's *logtransform!*, *normtransform!* and *decorrelate_by_mad!* methods. The nearly 100 lines of code dedicated to the quantification of morphological activity could be reduced to a one-liner and significantly accelerated thanks to parallelization and to the speed of Julia. This could be achieved via a single call to the *robust_morphological_perturbation_value* method, thus avoiding the definitions of the MCD computation, of the robust Mahalanobis distance, and of the permutation scheme, as well as the aggregation of the results.

Note that the Mahalanobis distance is defined between a single point and a distribution. The Hellinger distance generalizes this concept for two distributions, by incorporating estimators of location and scatter of two distributions, and is defined as follows:

$$H^2 = 1 - \frac{\det(S_1)^{\frac{1}{4}} \cdot \det(S_2)^{\frac{1}{4}}}{\det(S)^{\frac{1}{2}}} e^{-\frac{1}{8}(\mu_1 - \mu_2)^T S^{-1}(\mu_1 - \mu_2)},$$

with $S = (S_1 + S_2)/2$, where S_1 , μ_1 , S_2 and μ_2 are the covariance matrices and means of the distributions 1 and 2, respectively. As for the robust Mahalanobis distance, we can substitute the covariance matrices S_1 and S_2 and the centers μ_1 and μ_2 using the MCD estimators and thus define the robust Hellinger distance (RHD) that we used to quantify the distance between DMSO controls and each chemical perturbation. One requirement for the MCD computation, and therefore for using the RHD, is to have twice as many measurements per condition as dimensions. The filtering scheme described above results in some field-of-view profiles being discarded in many wells, yet most wells had more than eight valid fields of view. We thus chose to work in a four-dimensional space in order to characterize most treatments. To assess the statistical significance of these values, we conducted a permutation test by shuffling the label of the points (perturbation or control) and calculating again the RHD 5000 times, which formed a null distribution associated with an empirical P -value. As the statistical power of this test depends on the

number of permutations, an empirical P -value of zero corresponds to the case where no permutation led to a distance greater than the one actually observed between profiles and can be interpreted as an estimation of a P -value $< 1/5000$. To accelerate this process, the permutations were computed in parallel by distributing computations on 16 threads. After Benjamini–Hochberg false discovery rate (FDR) correction, we obtained a significance score coined the robust morphological perturbation value (RMPV) and defined all compounds with an RMPV < 0.1 , equivalent to an FDR cutoff of 10%, as morphological hits. Of note, the list of hits (Supplementary Table S1) was stable when doubling the number of permutations, showing that the process converged correctly.

To compare our results with other approaches that could be adopted with BioProfiling.jl, we quantified the Mahalanobis distance between the centroid (arithmetic mean of all points) and the centroid of the DMSO controls, either from unreduced profiles or after PCA transformation to two dimensions, which preserved 97.8% of the dataset variance. These distances were used in a permutation test as previously described to obtain FDR-corrected P -values describing how likely it is to observe such distances in the absence of a compound effect.

2.7 Morphological and network distances

We integrated morphological profile information with publicly available data about each compound. First, we collected mechanisms of action (MOAs) and molecular targets from the LINCS perturbation database (Stathias et al., 2020). We queried the Application Programming Interface (API) for exact name matches or removed pharmaceutical salts or chirality when necessary to find the correct compound. All annotations are presented in Supplementary Table S1. In total, 141 compounds had known targets and 112 were annotated with one or several MOAs. In particular, 23 MOAs were associated with 2 or more compounds and considered for downstream analysis. The largest changes induced for several MOAs were obtained using the *characteristic_features* method (Supplementary Table S3). To compare morphological profiles between pairs of MOAs, we projected the profiles of the 59 hit compounds in four dimensions using UMAP and computed pairwise RHDs as described above. The morphological distance between two MOAs was then defined as the average pairwise distance between compounds annotated to each MOA. We also obtained all human protein–protein interactions (PPIs) from the HIPPIE database (Alanis-Lobato et al., 2017), filtered out those with a confidence score below 0.63 (median of the score distribution), and assembled them into a PPI network. The conversion between gene symbols and ENTREZ identifiers of the targets was done with MyGene.info (Xin et al., 2016). We define the targets of an MOA as all known targets of the hit compounds associated with this MOA. We then assessed the network separation between the targets of each MOA using the s_{AB} score, which was previously found to be a good metric to study disease module and drug module separation (Caldera et al., 2019; Menche et al., 2015). The score is defined as

$$s_{AB} = \mu_{AB} - \frac{\mu_{AA} + \mu_{BB}}{2},$$

where μ_{AA} and μ_{BB} are the means of the minimum shortest network distances among the targets of MOA A and B, respectively, and μ_{AB} is the mean of the minimum shortest distance between the targets of MOA A and B.

2.8 Counting and classifying image artifacts

To quantify the prevalence of common imaging artifacts in HCS experiments, we visually inspected images from published studies deposited in the Image Data Resource (Breinig et al., 2015; Caldera et al., 2019; Gustafsdottir et al., 2013; Rohban et al., 2017; Williams et al., 2017). Depending on the study, we manually annotated either all four fields of view from the wells A1 to A6 and B1 to B6 (Breinig et al., 2015; Caldera et al., 2019) or all nine fields of view from the wells A1 to A3 and B1 to B3 (Gustafsdottir et al., 2013; Rohban et al., 2017). Artifacts included dye clots and

precipitations, cells not properly attached to the substrate, and other less frequent artifacts such as out-of-focus images or visible micro-well edges. Despite only covering a fraction of each plate and including parts of the well edge, where evaporation frequently leads to altered phenotypes (Bray *et al.*, 2016; Caicedo *et al.*, 2017), this sample demonstrates that there are many obstacles to overcome in HCS analyses. The most common artifacts only affected a restricted region of the image, suggesting that the unaffected parts of the images could be informative nonetheless, and motivating the extensive image filtering and quality controls performed in the respective studies. Given the overall abundance of such artifacts, however, we expect that a fraction of them will fail to be excluded and thus impact the image analysis and lead to outlier measurements. The issue was also present in the experiment we conducted, as artifact and outlier clusters were observed in the absence of filtering (Supplementary Fig. S1).

2.9 Profiling overexpression in Cell Painting experiments

We used a dataset from the Cell Painting Image Collection, a resource made publicly available under the CC0 1.0 license for the CytoData Hackathon 2018 and compiling several HCS experiments (Caicedo *et al.*, 2018). We retrieved CellProfiler measurements aggregated per well from two plates in an experiment characterizing the overexpression of 135 genes in A549 cells using the Cell Painting assay (Bray *et al.*, 2016), identified as ‘BBBC041-Caicedo’. From the untransformed measurements, we filtered out metadata and features related to object localization, excluded genetic targets with less than four replicates per plate, log-transformed the data and decorrelated the features by decreasing MAD as previously described. Finally, we reduced these profiles to two dimensions using UMAP, with the spread parameter set to 10 and other values left to default. The process was repeated independently for both plates, resulting in two sets of selected features and two distinct UMAP embeddings.

3 Results

3.1 Profiling chemical perturbations with BioProfiling.jl

We conducted and analyzed a chemical HCS experiment to study the morphological effect of small molecules in human osteosarcoma cells and demonstrate the applicability of BioProfiling.jl. In brief, we selected 311 compounds at a single concentration based on their morphological activity, and on their wide range of MOAs and disease associations. U-2-OS cells were seeded on top of drug plates, and were fixed and stained to display nuclei, F-actin and total protein. Fluorescence images were acquired at a 20× magnification (Fig. 2a). The images were analyzed with CellProfiler (McQuinn *et al.*, 2018) and morphological descriptors were measured for each cell. These measurements were imported in Julia and used to define an *Experiment* object to be processed with BioProfiling.jl (Fig. 2b). Two Jupyter notebooks enable the reproducibility of the following morphological profiling analysis (see Availability and implementation). First, filters are iteratively defined to identify cellular outliers based on extreme values. For instance, cells with unusually large cytoplasms compared to their nuclei were likely to be missegmented and therefore excluded (Fig. 2c). After aggregating the profiles per image and discarding the least informative features for characterizing chemical effects compared to DMSO controls, we reduced the dimensionality of the profiles to four dimensions using UMAP (McInnes *et al.*, 2018). The most discriminative measurements contained various descriptors of both nuclei and cytoplasm, as well as intensities of all dyes, suggesting that all aspects of morphology considered in our study were relevant (Supplementary Table S2). These features formed a morphological space in which the profiles of some compounds, such as Vinblastine (tubulin inhibitor) and Wiskostatin (actin polymerization inhibitor) but also Pentamidine (antifungal agent), were clustered away from images of DMSO treatment (Fig. 2d and Supplementary Fig. 2a). Using the dedicated methods for quantifying the significance of statistical distances implemented

in BioProfiling.jl, we identified 248 compounds with a significant morphological activity compared to DMSO controls in a plate seeded with 750 cells per well (Fig. 2e) with an FDR of 10%, coined morphological hits. In comparison, 242 hits were identified in a denser plate seeded with 1500 cells (Supplementary Fig. 2b). Of note, the seeding density had only a minor impact on whether compounds were identified as hits or not (Supplementary Fig. 2c). The hits on the two plates showed a large and highly significant overlap given the total number of tested compounds (Jaccard index of 0.78; χ^2 test of independence: $P = 1.5e-13$). This observation also held true for a more stringent FDR of 5% (Jaccard index of 0.73; χ^2 test of independence: $P = 9.0e-13$).

3.2 Investigating curation strategies

The curation of these profiles exemplified a particular set of methodological choices adapted to the specific experimental dataset at hand. BioProfiling.jl does not enforce any single approach and other options could be considered at all steps, from feature selection to dimensionality reduction and quantification of profile distances. As a comparison, we also used one of the other implemented distance metrics, namely the Mahalanobis distance between the center of a perturbation’s profiles and the reference profiles. When no dimensionality reduction was applied, we observed that all profiles were significantly distant from the DMSO. This reflects one consequence of the curse of dimensionality, namely that pairwise distances tend to be similar in high-dimensional spaces (Supplementary Fig. S3a). Using PCA in a strategy analogous to the mp-value (Hutz *et al.*, 2013), we obtained a space where the first axis explained the majority of the data variance, but still displayed some compound clustering (Supplementary Fig. S3b). When comparing the hits obtained with this approach to the initial hit list obtained with UMAP and the RHD (Supplementary Fig. S3c), we observed a partial but significant overlap (Jaccard index of 0.23; χ^2 test of independence: $P = 1.5e-13$). The hits obtained with PCA were only a subset of the hits initially obtained, suggesting that certain morphological changes can be better detected using nonlinear dimensionality reduction techniques. Overall, the choice of methodology has a considerable impact on which perturbations are identified as morphologically active, and consequently on all downstream analyses.

Note that BioProfiling.jl is not only versatile in respect to the profiling approaches, but also supports multiple experimental setups and data types, as any tabular data compatible with the common *DataFrame* structure can be used as input to define an *Experiment* object. We provide one additional example from a publicly available dataset using the Cell Painting assay (Bray *et al.*, 2016) to characterize overexpression constructs. By curating and representing profiles for two plates, we observe a visible clustering for some target genes (Supplementary Fig. S4a and b). Overall, comparing the RHD of each target to non-targeting controls correlates well across plates despite processing both plates independently (Supplementary Fig. 4c), which supports the robustness of the chosen approach.

3.3 Exploring MOAs of active compounds

We next went on to characterize the compounds with a strong morphological impact identified on the plate seeded with 750 cells per well using the approach relying on UMAP reduction and RHD. Among the wide range of MOAs covered by the library, 10 hit compounds were known Dopamine receptor antagonists, six hit compounds were annotated to Calcium channel blockers and six to Adrenergic receptor antagonists (Fig. 3a). In this experiment, some MOAs were likelier than others to induce morphological changes, often in accordance with their biological role. In particular, all Tubulin inhibitors caused cytoskeletal defects and were identified as hits. In contrast, only half of the Topoisomerase inhibitors, which modulate DNA replication and transcription and are more likely to impact cell shape indirectly, if at all, were found to modulate the morphology. We note the presence of many oncological and chemotherapeutic agents (PDGFR receptor inhibitors, Topoisomerase inhibitors, KIT inhibitors, Tubulin inhibitors) and neurological

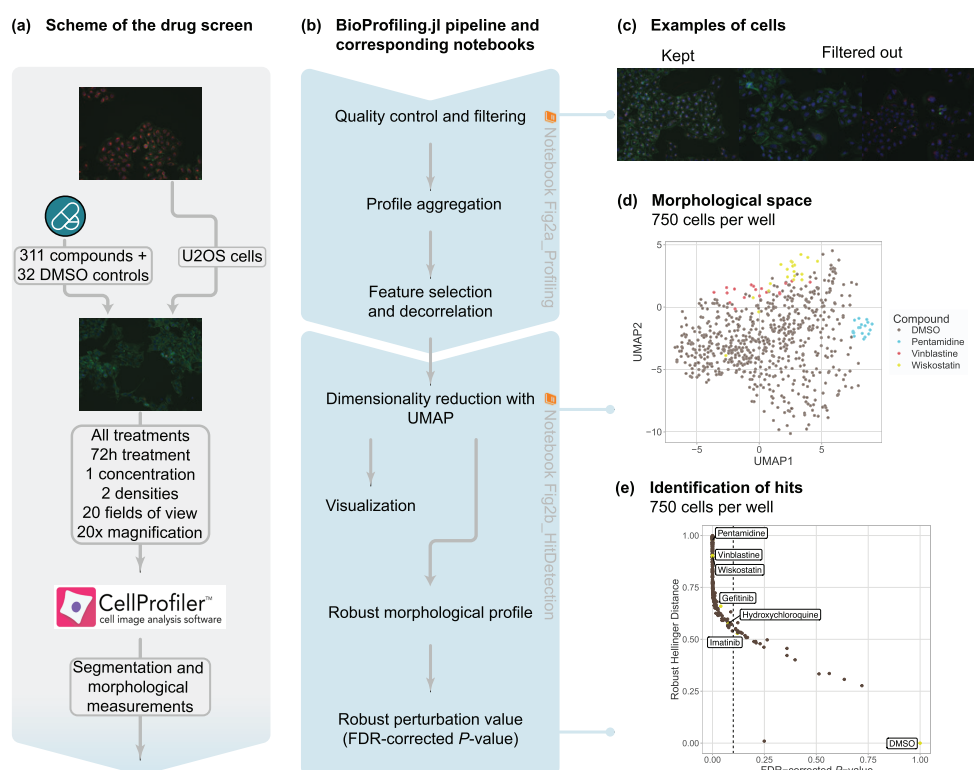


Fig. 2. Robust cellular profiling with BioProfiling.jl characterizes the morphological diversity induced by pharmacologically active compounds. (a) Experimental setup of the HCS experiment. Images are uncropped examples of untreated (top) and treated (middle) cells. (b) Computational workflow using BioProfiling.jl. Boxes are annotated with the name of the notebooks with which to reproduce the analyses. (c) Example of images displaying cells kept in the analysis (left) or problematic cells discarded by one of the quality-control filters (center, right). Cytoplasm and nucleus centers are marked with a white cross for each cell. (d) UMAP embedding preserving the cosine distance between the morphological profiles aggregated per field of view in the plate seeded with 750 cells/well. Two out of four dimensions are represented. (e) RHD and RMPV (FDR-corrected P -value) of each compound in the plate seeded with 750 cells/well compared to DMSO. Vertical dotted line indicates an FDR threshold of 0.1 and all compounds on its left are defined as morphological hits

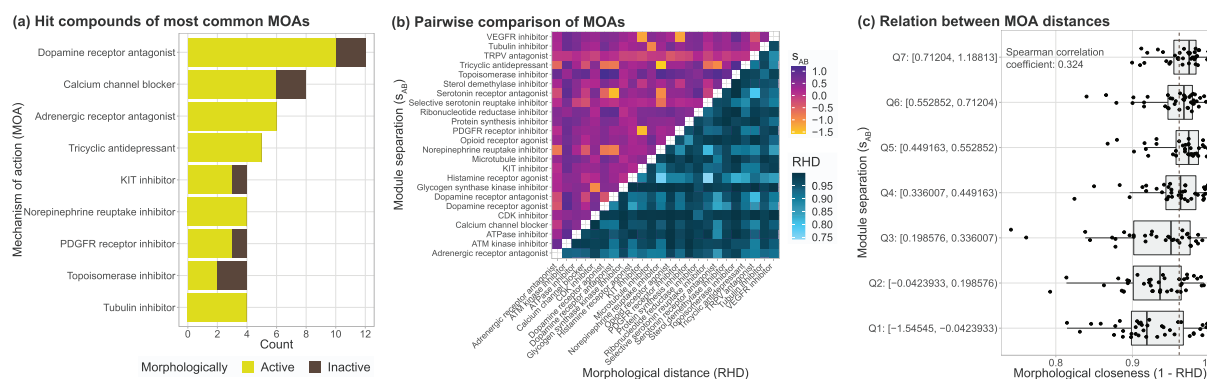


Fig. 3. Morphological profiling and data integration characterize compound MOAs. (a) Number of hits and total number of compounds for the most common MOAs in the chemical library. (b) Dissimilarity of the molecular targets on a PPI network (s_{AB} score, upper triangle) and of the morphological profiles (RHD, lower triangle) for the MOAs with at least two hit compounds. (c) Relation between drug module separation (bins of s_{AB} scores) and morphological distance (RHD)

drugs (Dopamine receptor antagonists, Tricyclic antidepressants, Norepinephrine reuptake inhibitors) among the morphological hits. Cell shape indeed plays an essential role in cancers, as cancerous cells are typically diagnosed by pathologists based on their morphology. Cell proliferation and several signaling pathways are also associated with cell geometry (Aragona et al., 2013; Dupont et al., 2011; Sero et al., 2015). Some compounds used to treat neurological disorders were also previously reported to induce morphological changes (Wawer et al., 2014), yet the mechanisms linking morphology and disease phenotype are still to be uncovered.

3.4 Integrating target properties and morphological profiles

We also compare effects between compounds to further exploit the richness of the morphological profiles. We quantified the similarity between the morphological impact of MOAs by aggregating the mean of the pairwise RHD between their respective hit profiles (Fig. 3b). While each MOA had a distinct signature, Glycogen synthase kinase inhibitors and CDK inhibitors were consistently distant from all other MOAs, hinting that modulation of kinase activity and cell signaling is likely to impact the cellular

morphology in broad and distinctive ways as opposed to inducing a particular cytoskeletal defect. The largest changes induced by compounds of these MOAs were impacting shape, intensities and distributions of multiple dyes (Supplementary Table S3). Of note, Kenpaulone is both a CDK inhibitor and a Glycogen synthase kinase inhibitor (Supplementary Table S1) which partly explains the observations shared for both MOAs.

While the morphological profiles are informative by themselves, they are best used by integrating additional information about the perturbations they describe. Here, we used available information on the targets of the compounds to contextualize their molecular environment within the PPI network. We quantified the network separation between the targets of different MOAs via the s_{AB} score, which was used previously to quantify the separation of disease and drug modules (Caldera *et al.*, 2019; Menche *et al.*, 2015). A positive score is associated with well-separated sets of nodes, whereas a negative score corresponds to an overlap. We found that all the strongest network overlaps between MOAs corresponded to shared drug classes. Tricyclic antidepressants and Norepinephrine reuptake inhibitors corresponded to non-selective monoamine reuptake inhibitors (ATC code N06AA). Serotonin receptor antagonists and Dopamine receptor antagonists both included Antipsychotics (ATC code N05A). PDGFR receptor inhibitors and KIT inhibitors were annotated to the exact same compound, Imatinib mesylate, Pazopanib and Sunitinib, which are all protein kinase inhibitors (ATC code L01E).

When comparing morphological profiles, Histamine receptor antagonists were close to many other MOAs and showed the most striking similarities with selective serotonin reuptake inhibitors and Norepinephrine reuptake inhibitors. All three affected primarily the cell shape (Supplementary Table S3). Of note, Histamine receptor antagonists displayed a consistent level of network similarity with all other MOAs. The s_{AB} values close to zero reflect in part the spread of their 12 molecular targets on the PPI network, suggesting that generic PPI alteration patterns may correspond to morphological effects that are distinctive, but not unique.

By comparing morphological distances to molecular network separation, we observed that overlapping target modules are associated with more similar morphological profiles (Fig. 3c). The effect does not fully explain morphological variability, which emphasizes the presence of intermediate regulatory processes between genotype and phenotype, and that the disruption of some biological processes is not detectable with general cell shape descriptors as experimental readouts. The quantification of morphological distances between profiles based on the UMAP-reduced space also means that part of the information contained in the original data is lost or distorted. Future development of methodologies leveraging the manifold learned by the UMAP method without the need for an embedding in Euclidean space will further alleviate this limitation. However, our results so far already confirm that there is a general association between the PPI network neighborhood targeted by a compound and their morphological outcome. This could be further explored to systematically link cellular morphology to function in health and disease.

4 Discussion

HCS experiments offer a scalable and cost-efficient way to assess multiple conditions in a single experiment with a rich cellular readout. Assembling morphological profiles to describe these experimental conditions is thus essential and requires dedicated tools for data curation, feature selection, quality control, visualization and quantification of morphologically active perturbations. We implemented these tools in a single open-source software with intuitive and flexible data structures and syntax. We demonstrated by a concrete use case how BioProfiling.jl enables new research and allows the exploration of changes in cellular morphology by easing the analysis of large high-content imaging screens.

As Julia is an efficient programming language and allows parallelization of the computations, BioProfiling.jl can process large

datasets in a performant manner. The biggest limitation for analyzing large experiments at the single cell level is currently the memory usage, as the full set of morphological measurements needs to be loaded, which can be an issue on personal computers. This may be improved in the future by using lazy loading and allowing the user to process the data by batches. In regard to profile interpretability, BioProfiling.jl can help identify which features are varying the most, rank features by absolute fold-change when comparing two conditions, highlight correlated measurements and format the data for other tools, for instance to represent the typical cell morphology in a particular condition (Khawatmi *et al.*, 2021; Sailem *et al.*, 2015). Of note, BioProfiling.jl offers a systematic way to define filters for data curation and feature selection. This simplifies the automated definition of these steps and could contribute toward the future development of data-driven feature engineering and machine-learning-powered artifact removal techniques to further streamline the process of morphological profiling.

BioProfiling.jl expands the existing landscape of resources available for biological data analysis, as illustrated in our application where we processed morphological measurements so that they can be integrated with PPIs as well as chemical annotations. The library contributes to the growing package ecosystem for bioinformatics in Julia (Roesch *et al.*, 2021) which ensures that the morphological profiling analyses can be combined in larger projects together with other tasks ranging from sequencing to systems biology (Greener *et al.*, 2020; Heirendt *et al.*, 2017; Zea *et al.*, 2016), and other libraries are conveniently available to integrate these different data types (Zakeri *et al.*, 2018). Julia's interoperability with other programming languages also makes the onboarding easy for users with prior programming experience who, for instance, might prefer to perform certain tasks in R or Python. This is demonstrated in the provided Jupyter notebooks, with all plots being generated using R's ggplot2 library (Wickham, 2016) and with the computation of the MCD estimators for robust statistical distances which relies on R's robust-base package, which in turn calls efficient Fortran routines.

Despite being initially designed and extensively tested for morphological profiling, the ability of BioProfiling.jl to handle large high-dimensional datasets and provide dedicated robust normalization and comparison methods could also be leveraged for other data analyses such as single cell transcriptomics or metabolomics experiments, which also require the curation and transformation of data in tabular format.

To provide an exemplary full use case of the BioProfiling.jl package, we conducted and analyzed a chemical high-content imaging screen for characterizing the effect of small molecules across diverse MOAs. The compounds used in the screen were selected to cover a wide range of morphological activity. While large-scale, hypothesis-free screening of small molecules can offer an unbiased view of the compound types that affect cellular morphology (Bryce *et al.*, 2019; Wawer *et al.*, 2014), our library design enabled us to observe significant changes induced by more than three quarters of the used compounds. At the same time, the focused library design limits the interpretation of MOA enrichment among hits. We observed both commonalities and differences in the effects induced by different MOAs, which alter cellular morphology via different molecular changes, involving cytoskeleton, nucleus and protein relocation (Supplementary Tables S2 and S3). The corresponding morphological profiles were further integrated with the information available about the PPI network properties of the compound targets, which proved to offer complementary views of compound effects and emphasized the role HCS could play in unraveling the relationship between cellular morphology and function.

Acknowledgements

We thank Raphael Bednarsky for discussion and feedback on the library and Daniel Malzl for his feedback on the manuscript.

Funding

This work was supported by the Vienna Science and Technology Fund (WWTF) through projects VRG15-005 (to J.M.) and LS16-060 (to J.M. and L.D.) and by the CNRS (International Research Project SysTact to L.D.). C.W.F. was supported by a DOC-fellowship of the Austrian Academy of Sciences: 25525.

Conflict of Interest: none declared.

References

- Alanis-Lobato, G. *et al.* (2017) HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.*, **45**, D408–D414.
- Ando, D.M. *et al.* (2017) Improving phenotypic measurements in high-content imaging screens. *bioRxiv*, <https://doi.org/10.1101/161422>.
- Aragona, M. *et al.* (2013) A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell*, **154**, 1047–1059.
- Becker, T. *et al.* (2021) cytominer: methods for image-based cell profiling. <https://github.com/cytomining/cytominer>.
- Bezanson, J. *et al.* (2017) Julia: a fresh approach to numerical computing. *SIAM Rev.*, **59**, 65–98.
- Billmann, M. *et al.* (2016) A genetic interaction map of cell cycle regulators. *Mol. Biol. Cell*, **27**, 1397–1407.
- Bougen-Zhukov, N. *et al.* (2017) Large-scale image-based screening and profiling of cellular phenotypes. *Cytometry Part A*, **91**, 115–125.
- Boutros, M. *et al.* (2006) Analysis of cell-based RNAi screens. *Genome Biol.*, **7**, R66.
- Boutros, M. *et al.* (2015) Microscopy-based high-content screening. *Cell*, **163**, 1314–1325.
- Bray, M.-A. *et al.* (2016) Cell Painting: a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.*, **11**, 1757–1774.
- Breinig, M. *et al.* (2015) A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.*, **11**, 846–846.
- Bryce, N.S. *et al.* (2019) High-content imaging of unbiased chemical perturbations reveals that the phenotypic plasticity of the actin cytoskeleton is constrained. *Cell Syst.*, **9**, 496–507.E5.
- Cabana, E. *et al.* (2019) Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Stat. Pap.*, **62**, 1583–1609.
- Caicedo, J.C. *et al.* (2017) Data-analysis strategies for image-based cell profiling. *Nat. Methods*, **14**, 849–863.
- Caicedo, J.C. *et al.* (2018) Cell Painting image collection. <https://registry.opendata.aws/cell-painting-image-collection>.
- Caldera, M. *et al.* (2019) Mapping the perturbome network of cellular perturbations. *Nat. Commun.*, **10**, 5140.
- Chandrasekaran, S.N. *et al.* (2020) Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat. Rev. Drug Discov.*, **20**, 145–159.
- Dupont, S. *et al.* (2011) Role of YAP/TAZ in mechanotransduction. *Nature*, **474**, 179–183.
- Eling, N. *et al.* (2021) cytomapper: an R/Bioconductor package for visualization of highly multiplexed imaging data. *Bioinformatics*, **36**, 5706–5708.
- Fischer, B. *et al.* (2015) A map of directional genetic interactions in a metazoan cell. *eLife*, **4**, e05464.
- German, Y. *et al.* (2021) Morphological profiling of human T and NK lymphocytes by high-content cell imaging. *Cell Rep.*, **36**, 109318.
- Greener, J.G. *et al.* (2020) BioStructures.jl: read, write and manipulate macromolecular structures in Julia. *Bioinformatics*, **36**, 4206–4207.
- Gustafsdottir, S. *et al.* (2013) Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One*, **8**, e80999.
- Heigwer, F. *et al.* (2018) Time-resolved mapping of genetic interactions to model rewiring of signaling pathways. *eLife*, **7**, e40174.
- Heirendt, L. *et al.* (2017) DistributedFBA.jl: high-level, high-performance flux balance analysis in Julia. *Bioinformatics*, **33**, 1421–1423.
- Hutz, J.E. *et al.* (2013) The multidimensional perturbation value. *J. Biomol. Screen.*, **18**, 367–377.
- Jones, T.R. *et al.* (2005) Voronoi-based segmentation of cells on image manifolds. In: *Proceedings of the First International Conference on Computer Vision for Biomedical Image Applications, CVBIA'05*. Springer-Verlag, Berlin, Heidelberg, pp. 535–543.
- Jones, T.R. *et al.* (2008) CellProfiler analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinform.*, **9**, 482.
- Khawatmi, M. *et al.* (2021) ShapoGraphy: a glyph-oriented visualization approach for creating pictorial representations of bioimaging data. *bioRxiv*.
- Ljosa, V. *et al.* (2013) Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J. Biomol. Screen.*, **18**, 1321–1329.
- Lu, A.X. *et al.* (2019) Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.*, **15**, e1007348.
- McInnes, L. *et al.* (2018) UMAP: uniform manifold approximation and projection. *J. Open Source Softw.*, **3**, 861.
- McQuinn, C. *et al.* (2018) CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.*, **16**, e2005970.
- Menche, J. *et al.* (2015) Uncovering disease-disease relationships through the incomplete interactome. *Science*, **347**, 1257601.
- Pau, G. *et al.* (2010) EBIImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, **26**, 979–981.
- Pawlowski, N. *et al.* (2016) Automating morphological profiling with generic deep convolutional networks. *bioRxiv*, <https://doi.org/10.1101/085118>.
- Roesch, E. *et al.* (2021) Julia for biologists. <https://arxiv.org/abs/2109.09973>.
- Rohban, M.H. *et al.* (2017) Systematic morphological profiling of human gene and allele function via Cell Painting. *eLife*, **6**, e24060.
- Rousseeuw, P.J. and van Driessen, K. (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**, 212–223.
- Sailem, H.Z. *et al.* (2015) Visualizing cellular imaging data using PhenoPlot. *Nat. Commun.*, **6**, 5825.
- Scheeder, C. *et al.* (2018) Machine learning and image-based profiling in drug discovery. *Curr. Opin. Syst. Biol.*, **10**, 43–52.
- Sero, J.E. *et al.* (2015) Cell shape and the microenvironment regulate nuclear translocation of NF- κ B in breast epithelial and tumor cells. *Mol. Syst. Biol.*, **11**, 790.
- Simm, J. *et al.* (2018) Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.*, **25**, 611–618.e3.
- Snijder, B. *et al.* (2017) Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study. *Lancet Haematol.*, **4**, e595–e606.
- Stathias, V. *et al.* (2020) LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Res.*, **48**, D431–D439.
- Wawer, M.J. *et al.* (2014) Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. USA*, **111**, 10911–10916.
- Way, G.P. *et al.* (2021a) Predicting cell health phenotypes using image-based morphology profiling. *Mol. Biol. Cell.*, **32**, 823–1005.
- Way, G.P. *et al.* (2021b) Pycytominer: data processing functions for profiling perturbations. <https://github.com/cytomining/pycytominer>.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, NY.
- Williams, E. *et al.* (2017) Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods*, **14**, 775–781.
- Xin, J. *et al.* (2016) High-performance web services for querying gene and variant annotation. *Genome Biol.*, **17**, 91.
- Zakeri, P. *et al.* (2018) Gene prioritization using Bayesian matrix factorization with genomic and phenotypic side information. *Bioinformatics*, **34**, i447–i456.
- Zea, D.J. *et al.* (2016) MIToS.jl: mutual information tools for protein sequence analysis in the Julia language. *Bioinformatics*, **33**, 564–565.
- Ziegler, S. *et al.* (2021) Morphological profiling of small molecules. *Cell Chem. Biol.*, **28**, 300–319.

3.2 Morphological profiling of human T and NK lymphocytes by high-content cell imaging

Yolla German*, **Loan Vulliard***, Anton Kamnev, Laurène Pfajfer, Jakob Huemer, Anna-Katharina Mautner, Aude Rubio, Artem Kalinichenko, Kaan Boztug, Audrey Ferrand, Jörg Menche and Loïc Dupré.

Published in Cell Reports. Volume 36, article number 109318. Publication date: July 6, 2021.
DOI: <https://doi.org/10.1016/j.celrep.2021.109318>.

* These authors contributed equally

The human immune system relies on the synergy between multiple specialized cell types to protect us from external threats such as pathogens, and internal defects such as cancerous cells. The immune system is usually divided into two arms, namely the innate response, forming the first line of defense and the adaptive response, slower but more targeted. Although this conceptual separation is convenient, the identification of interconnections and cell types at their interface has been blurring the line between the two response types (Dranoff, 2004). In both arms, lytic granules are mediating the elimination of infected cells. They contain perforin monomers, which polymerize to form pores at the surface of the target cell, and granzymes, which contribute to induce programmed cell death (Osińska et al., 2014). These granules are produced and released by cytotoxic lymphocytes, whose most prominent examples are natural killer (NK) cells and CD8+ T cells for the innate and adaptive arms of the immune system, respectively. In brief, target cells typically present non-self peptides via the major histocompatibility complex (MHC) of type I, which get recognized by receptors harbored by the cytotoxic T cells. However, some cancerous or virus-infected cells evade detection by not expressing MHC molecules. NK cells precisely detect and repress this behavior (Campbell & Hasegawa, 2013).

The immunological synapse (IS) is the structure at the interface between a lymphocyte and the cell being targeted. The dynamic architecture of the IS relies on actin re-organization to be functional and to allow the release of the lytic granules. Because of this direct link between spatial organization and cytotoxic activity, microscopy is an approach of choice to study the IS. So far, microscopy-based immunological studies were typically constrained in their scale and conducted in microplates with few wells. Experimentalists only selected and imaged a limited number of cells, allowing for the comparison of a restricted set of morphological parameters. In contrast to this approach, the research group of Dr. Loïc Dupré aimed to leverage the throughput and flexibility offered by HCI.

In this article, we demonstrate that this experimental approach is ideal for assessing the morphology of cytotoxic lymphocytes. Here, the microplates were coated with a stimulatory medium to make the cells adherent and additionally to mimic the formation of the IS. We show that HCI enables the characterization of subtle changes in morphology and the discrimination

between different alterations of the actin cytoskeletons, either induced chemically or resulting from actinopathies. The flexibility of the approach allowed us to study multiple experimental and biological conditions, with both T and NK lymphocytes, represented by cell lines, primary and patient-derived cells. Moreover, by sharing openly the complete data and computational tools generated, this study constitutes a valuable resource to cell biologists and immunologists interested in perturbations of the cytoskeleton or aiming to characterize other conditions or actinopathies based on HCI.

This project was initiated by Dr. Yolla German and Dr. Loïc Dupré, who planned and performed the experiments. In the following publication, I designed and conducted the morphological profiling in all experiments and carried out computational analyses, including data visualization and characterization of morphological changes using random forest classifiers and robust statistics. The article was written by Dr. Yolla German, Dr. Loïc Dupré and myself with contributions from all co-authors. See the section of the paper entitled “AUTHOR CONTRIBUTIONS” for more information.

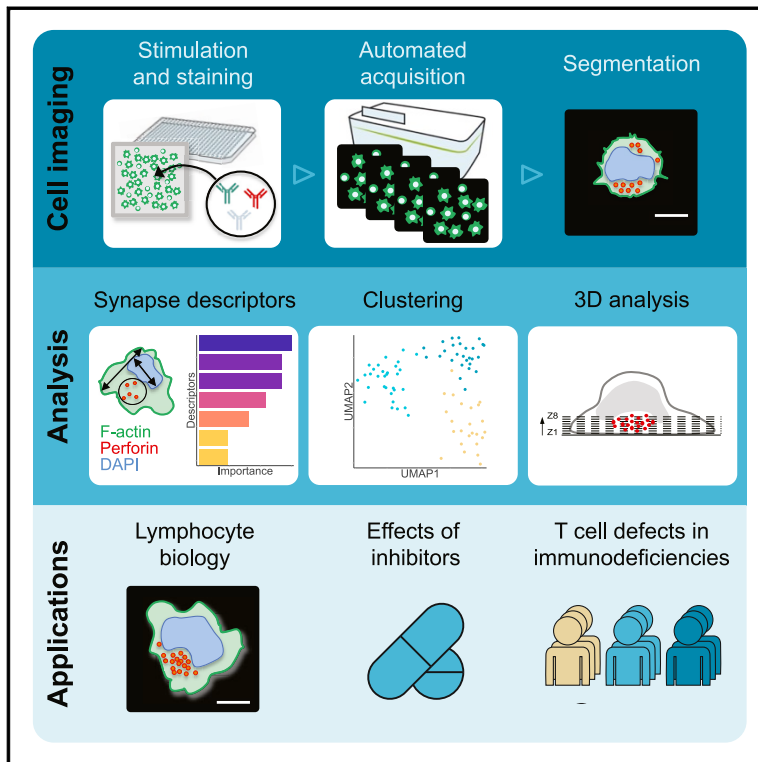
Of note, a complementary view of this project, providing additional insight regarding the experimental setup and biological context, is proposed in the doctoral thesis of Dr. Yolla German, completed at the Université Toulouse III - Paul Sabatier (German, 2020).

Supplementary material can be found below in the appendix chapter.

Cell Reports

Morphological profiling of human T and NK lymphocytes by high-content cell imaging

Graphical abstract



Authors

Yolla German, Loan Vulliard, Anton Kamnev, ..., Audrey Ferrand, Jörg Menche, Loïc Dupré

Correspondence

loic.dupre@inserm.fr

In brief

German et al. provide a workflow to study activation of cytotoxic T and NK lymphocytes by high-content confocal microscopy. Images are analyzed to characterize morphological variation in immunological synapses induced by actin compounds in primary cells as well as in CD8⁺ T cells derived from ARPC1B- and WASP-deficient individuals.

Highlights

- Cytotoxic lymphocytes can be profiled in depth using high-content imaging
- Morphological changes are identified with dedicated analytical tools
- Distinct immunological synapse alterations in ARPC1B- and WASP-deficient T cells
- 3D distribution of lytic granules is impaired in ARPC1B-deficient CD8⁺ T cells



German et al., 2021, Cell Reports 36, 109318
 July 6, 2021 © 2021 The Author(s).
<https://doi.org/10.1016/j.celrep.2021.109318>

Resource

Morphological profiling of human T and NK lymphocytes by high-content cell imaging

Yolla German,^{1,2,11} Loan Vulliand,^{3,4,11} Anton Kamnev,^{2,5} Laurène Pfajfer,^{1,2} Jakob Huemer,^{2,6} Anna-Katharina Mautner,^{2,5} Aude Rubio,⁷ Artem Kalinichenko,^{2,6} Kaan Boztug,^{2,3,6,8,9} Audrey Ferrand,⁷ Jörg Menche,^{3,4,10} and Loïc Dupré^{1,2,5,12,*}

¹Toulouse Institute for Infectious and Inflammatory Diseases (INFINITY), INSERM UMR1291, CNRS UMR5051, Toulouse III Paul Sabatier University, Toulouse, France

²Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD), Vienna, Austria

³CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

⁴Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Vienna, Austria

⁵Department of Dermatology, Medical University of Vienna, Vienna, Austria

⁶St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria

⁷IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS, 31024 Toulouse, France

⁸Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria

⁹St. Anna Children's Hospital, Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria

¹⁰Faculty of Mathematics, University of Vienna, Vienna, Austria

¹¹These authors contributed equally

¹²Lead contact

*Correspondence: loic.dupre@inserm.fr

<https://doi.org/10.1016/j.celrep.2021.109318>

SUMMARY

The immunological synapse is a complex structure that decodes stimulatory signals into adapted lymphocyte responses. It is a unique window to monitor lymphocyte activity because of development of systematic quantitative approaches. Here we demonstrate the applicability of high-content imaging to human T and natural killer (NK) cells and develop a pipeline for unbiased analysis of high-definition morphological profiles. Our approach reveals how distinct facets of actin cytoskeleton remodeling shape immunological synapse architecture and affect lytic granule positioning. Morphological profiling of CD8⁺ T cells from immunodeficient individuals allows discrimination of the roles of the ARP2/3 subunit ARPC1B and the ARP2/3 activator Wiskott-Aldrich syndrome protein (WASP) in immunological synapse assembly. Single-cell analysis further identifies uncoupling of lytic granules and F-actin radial distribution in ARPC1B-deficient lymphocytes. Our study provides a foundation for development of morphological profiling as a scalable approach to monitor primary lymphocyte responsiveness and to identify complex aspects of lymphocyte micro-architecture.

INTRODUCTION

Recent advances in high-content imaging (HCI) allow morphological profiling of cell populations at a rich level of detail, providing an integrative readout for complex biological events. HCI is particularly suitable for systematic analysis of cellular phenotypes. It is employed widely in cancer and toxicology research on adherent cell lines to evaluate cellular fitness and viability upon drug treatment or genetic alterations (de Groot et al., 2018; Joshi and Lee, 2015; Papakonstantinou and O'Brien, 2014; Priestley et al., 2019; Wu and Li, 2018). HCI is also a powerful approach for identification of novel systemic properties of cells (Bryce et al., 2019; Caldera et al., 2019; Heigwer et al., 2018; Mattiazzi Usaj et al., 2020). To the best of our knowledge, HCI has not yet been applied to study immune cell populations, which are traditionally studied by flow cytometry. Because these highly specialized cells adopt distinctive morphologies and structures to sustain their various immuno-surveillance tasks, HCI would be highly relevant to monitor immune cell shape re-

modeling upon activation. In particular, the immunological synapse (IS) is a complex structure that sets lymphocyte activation and function during encounters with antigen-presenting cells and target cells. Low-throughput microscopy approaches have revealed that the IS is characterized by a symmetrical architecture consisting of concentric rings of F-actin and integrins, whereas the antigen receptors occupy a central position (Grouk et al., 1999; Monks et al., 1998). The lymphocyte spreading associated with IS assembly and organization relies on dynamic rearrangements of the actin cytoskeleton (Dustin and Long, 2010). In cytotoxic lymphocytes, including CD8⁺ T cells and natural killer (NK) cells, the IS is particularly important because it sustains polarized delivery of cytolytic molecules, such as perforin and granzymes, toward target cells (Somersalo et al., 2004). The study of IS organization has remained low throughput and has been restricted to analysis of a limited number of morphological measurements. A more systematic and in-depth assessment of the IS would leverage this structure as a pivotal readout for characterization of lymphocyte activation.



In this study, we report implementation of a scalable HCI approach combining confocal imaging of human T and NK cells stimulated over 2D surfaces functionalized with ICAM-1 and stimulatory antibodies (Abs), and high-dimensional profiling and clustering of IS morphologies. Comparison of lymphocytes treated with drugs affecting different facets of actin cytoskeleton remodeling highlights a variety of morphological alterations and effects on lytic granule distribution and degranulation. Application of our HCI pipeline to lymphocytes isolated from human blood reveals distinct morphological profiles in individual healthy donors. Furthermore, our method allows discrimination of synapse defects in untransformed CD8⁺ T cells from individuals with related deficiencies in the actin regulators ARPC1B and WASP, illustrating its potential to identify disease-specific morphological signatures. Single-cell analysis points to a loss of relationship between actin radial distribution and lytic granule dispersion as a possible explanation for the reduced cytotoxic activity of ARPC1B-deficient CD8⁺ T cells. This inspired complementary time-lapse recording of actin and lytic granule dynamics, indicating unstable positioning of the lytic granule pool in ARPC1B-deficient CD8⁺ T cells. Finally, we detail our experimental and computational methods and provide a reproducible analysis environment to enable their application to other study models.

RESULTS

Morphological profiles of T and NK cell ISs

To systematically analyze the morphological profile of T and NK cells, we designed a custom HCI workflow consisting of seeding cells on stimulatory surfaces in microwells of 96- or 384-well plates with subsequent fixation and staining for nuclei, the actin cytoskeleton, and perforin or LFA-1 with combinations of fluorescent dyes and Abs. Confocal images were acquired on an Opera Phenix high-content screening system equipped with a water immersion 40× objective and 4 scientific complementary metal-oxide-semiconductor (sCMOS) cameras. Images were acquired along 8 Z-planes with a 0.5-μm interval starting from the cell-substrate interface and were analyzed with CellProfiler (McQuinn et al., 2018) to automatically segment individual cells and extract quantitative morphological descriptors based on the different fluorescent markers (Figure 1A). As proof of concept, we applied our approach to NK-92 and Jurkat cells, two human cell lines commonly used as models for NK cells and T cells, respectively. Cell morphologies were compared upon interaction with a poly-L-lysine (PLL)-coated or stimulatory surface with the LFA-1 ligand ICAM-1 and stimulatory Ab evoking IS assembly. Upon co-stimulation with ICAM-1 and anti-NKp30/NKp46 Ab, NK-92 cells spread, emitted F-actin-rich peripheral pseudopodia, and gathered perforin-containing granules toward the center of the cell-substrate interface, as indicated by their detection in proximity of the cell-substrate interface (Figure 1B; Figure S1A). These observations are in line with the hallmarks of the IS in cytotoxic lymphocytes (Dieckmann et al., 2016; Dustin and Long, 2010), validating our high-throughput stimulation and staining procedure. We first assembled a curated set of descriptors to serve as the basis for morphological profiling. Based on literature describing the IS and reporting polarization of F-actin and lytic granules in NK cells (Krzewski and Coligan, 2012; Mace et al., 2012, 2014), we in-

tegrated morphological descriptors for F-actin and perforin staining and completed them with nuclear measurements via DAPI staining (Figure 1C; Table S1). The selected descriptors were extracted as mean values per field of view averaged across 3 experiments. The increases in F-actin intensity and cell area were prominent features of the stimulation compared with the PLL condition. Furthermore, the number of perforin-containing granules detected at the cell-substrate interface increased upon stimulation, which is indicative of their enrichment toward the IS. We detected relative spreading of the area covered by lytic granules, supporting the notion of multiple docking domains at the synapse (Brown et al., 2011). Notably, our analysis also highlights an increase in nucleus area as a typical feature of the IS in stimulated NK-92 cells. The nucleus area increased noticeably along with F-actin intensity when assessed across three experiments (Figure 1D), suggesting that cortical actin is associated with mechanical tensions that may flatten the nucleus. Of note, the absolute values for F-actin intensity were higher in one of the 3 experiments, possibly resulting from differences in staining quality. This indicates that absolute values for staining intensities across experiments should be considered with caution. Consequently, all further comparisons of samples and treatments were done within the same staining/acquisition batches and within the same plates. To ensure that no morphological heterogeneity across cells was overlooked by comparing mean values per field of view, F-actin intensity was assessed on a per-cell basis (Figure 1E). The unimodal increase in F-actin intensity driven by stimulation of NK-92 cells indicates a relatively homogeneous IS assembly, which validates our approach.

In parallel, the HCI workflow was applied to Jurkat cells, which were stimulated with ICAM-1 and anti-CD3 Abs. We selected, in this instance, 12 measurements to constitute our curated set of morphological descriptors, which pertained to F-actin, LFA-1, and DAPI staining to monitor hallmarks of the T cell IS (Dustin and Long, 2010). Because Jurkat cells do not express perforin, the LFA-1 integrin was stained instead. Compared with the neutral PLL surface, ICAM-1/anti-CD3 Ab stimulation led to cell spreading, assembly of a peripheral F-actin ring-like structure, and distribution of the integrin LFA-1 as a dense array of clusters at the cell-substrate interface (Figure 1F; Figure S1B), which are characteristic for the IS (Dustin and Long, 2010; Houmadi et al., 2018). Our quantification over multiple fields showed that, similar to NK-92 cells, F-actin intensity and cell area are prominent features of the Jurkat cell IS (Figure 1G; Table S2). Moreover, LFA-1 intensity and LFA-1 area also increased upon stimulation, and we noted that F-actin and LFA-1 intensities correlated in individual fields of view with a Pearson correlation coefficient of 0.50 (Figure 1H). At the single-cell level, F-actin clearly increased in response to ICAM-1 and anti-CD3 Ab stimulation despite noticeable heterogeneity (Figure 1I). The data collected on assembly of the IS in NK-92 and Jurkat cells reveal commonalities, such as a raise in F-actin intensity and cell spreading, whereas Jurkat cells, but not NK-92 cells, became rounder upon co-stimulation of LFA-1 and T cell receptors (TCR). Furthermore, although NK-92 cells flattened their nuclei, this effect was not observed in Jurkat cells, indicating a distinct cell spreading behavior in each cell line. Such differences among the 2 cellular models might be due to distinct IS architectures or

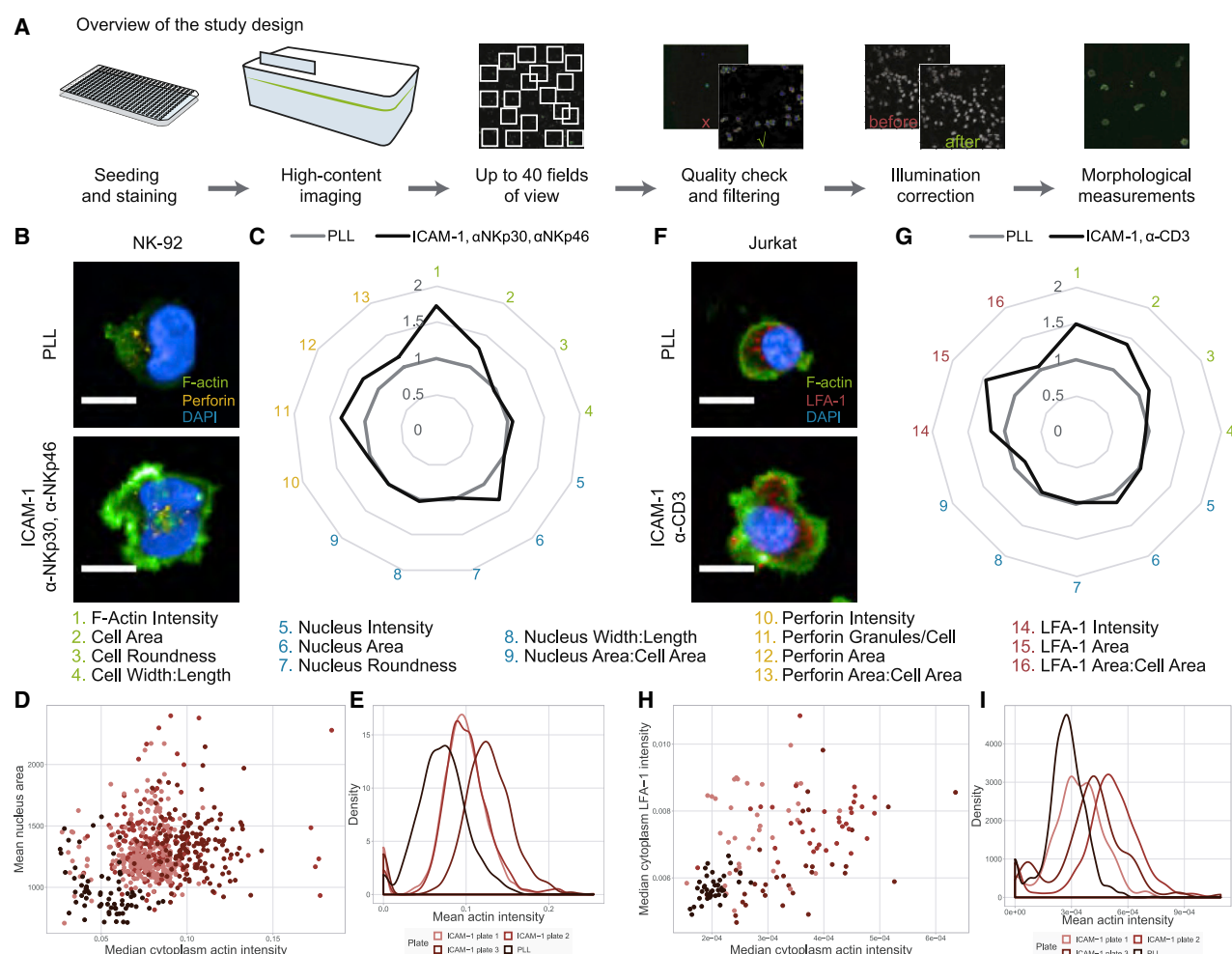


Figure 1. HCI of the IS in lymphocytic cell lines

(A) Schematic of the HCI pipeline.

(B) Representative images of NK-92 cells seeded on poly-L-lysine (PLL) or ICAM-1 plus anti-NKp30/NKp46 Abs, stained for F-actin (green), perforin (yellow), and the nucleus (DAPI). Scale bars, 10 μ m.

(C) Selected IS features analyzed as fold change of ICAM-1 plus anti-NKp30/NKp46 Abs over PLL. The data represent the mean of three separate experiments ($n = 933$ –5,860 cells).

(D) Mean nucleus area in pixels and median F-actin intensity per image across PLL and ICAM-1 plus anti-NKp30/NKp46 Ab conditions.

(E) F-actin intensity distribution per cell across PLL and ICAM-1 plus anti-NKp30/NKp46 Ab conditions.

(F) Representative images of Jurkat cells on PLL or ICAM-1 plus anti-CD3 Abs, stained for F-actin (green), LFA-1 (red), and the nucleus (DAPI). Scale bars, 10 μ m.

(G) Selected IS features analyzed as fold change of ICAM-1 plus anti-CD3 Abs over PLL. The data represent the mean of triplicates ($n = 125$ –940 cells).

(H) Median F-actin and LFA-1 intensity per image across PLL and ICAM-1 plus anti-CD3 Ab conditions.

(I) F-actin intensity distribution per cell across PLL and ICAM-1 plus anti-CD3 Ab conditions.

See also [Figure S1](#) and [Tables S1](#) and [S2](#).

different kinetics of the IS assembly, as highlighted recently by a comparison of primary T and NK cells ([Carisey et al., 2018](#)). Overall, our data validate the reliability and power of our HCI pipeline for automated morphological profiling of stimulated lymphocytes with high spatial resolution.

Cytoskeleton drugs induce distinct alterations of the NK cell IS

In line with the role of actin remodeling in sustaining IS assembly ([Dustin and Cooper, 2000](#)), our data point to the rise in F-actin

content as a prominent feature of this process. To gain a systematic view of how distinct facets of actin dynamics, including polymerization, coupling to myosin, and branching or elongation of filaments, might contribute to assembly of ISs in cytotoxic lymphocytes, we performed morphological profiling of NK-92 cells stimulated with ICAM-1 and anti-NKp30/NKp46 Abs upon treatment with three concentrations of the drugs latrunculin B, jasplakinolide, blebbistatin, Y-27632, CK-869, wiskostatin, and SMIFH2. To compare morphological alterations, we used curated morphological descriptors identified previously ([Figures 2A](#) and

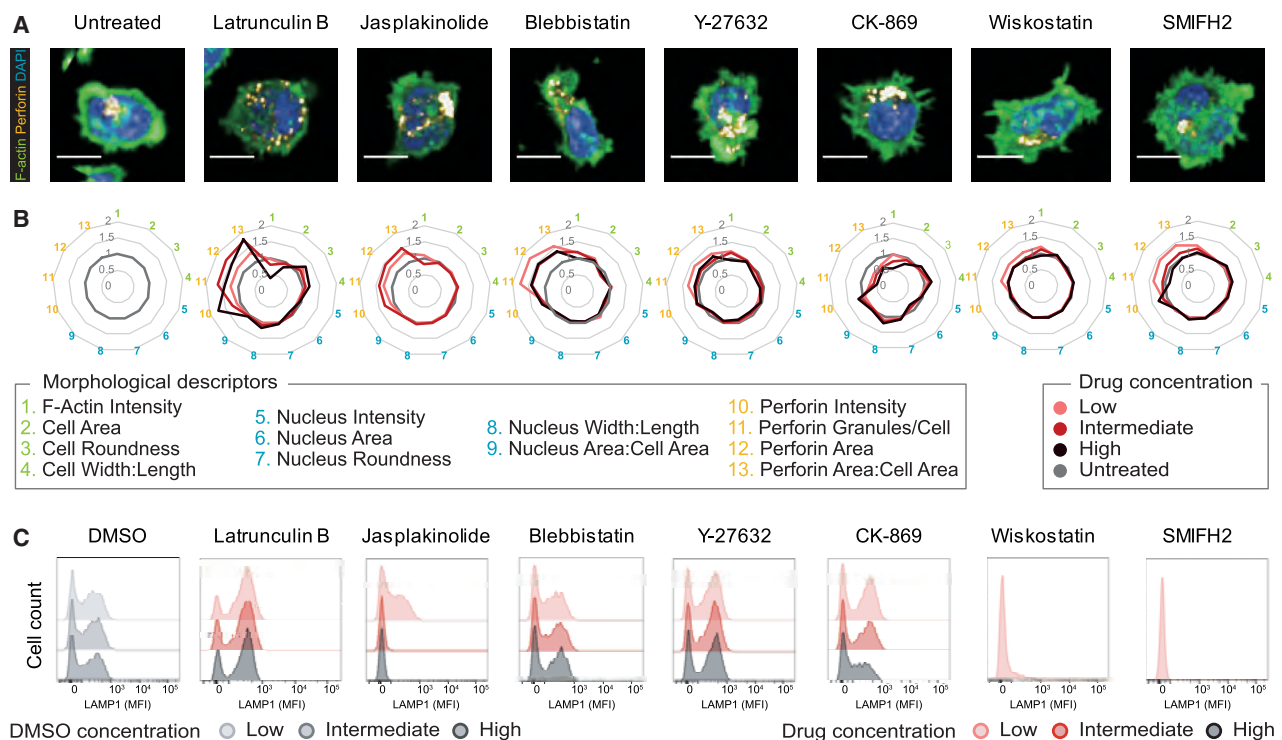


Figure 2. Comparative analysis of the effect of actin-targeting drugs on lytic granule distribution and exocytosis in NK-92 cells

(A and B) Representative images (A) and graphs (B) representing the fold change of IS descriptors for NK-92 cells seeded on ICAM-1, anti-NKp30, and anti-NKp46, stained for F-actin (green), perforin granules (yellow), and nuclei (blue) and treated with three concentrations of latrunculin B, jasplakinolide, blebbistatin, Y-27632, CK-869, wiskostatin, and SMIFH2 with respect to the untreated control. The data represent the mean of triplicates for each concentration (1,425–5,541 cells). Scale bars, 10 μ m.

(C) Evaluation of LAMP-1 expression on NK-92 cells upon stimulation and treatment with the indicated drugs.

2B). This analysis was complemented with measurement of LAMP-1 surface exposure to assess the effects of the tested drugs on lytic granule exocytosis (Figure 2C; Figure S2). As expected, inhibition of actin polymerization by latrunculin B led to a concentration-dependent decrease in F-actin intensity (Figures 2A and 2B), validating our approach. Although cell spreading was not fully impeded, as revealed by only a minor reduction in cell area, cell roundness increased, possibly because of a reduction in peripheral actin-rich protrusions. Latrunculin B treatment elicited an increase in the number and area of perforin granules. In parallel, the treated cells exhibited increased LAMP-1 surface exposure upon stimulation (Figure 2C). This suggests that the loss of F-actin density evoked by latrunculin B treatment impaired the filter function of the F-actin cortical network (Carisey et al., 2018; Mace et al., 2012), resulting in uncontrolled and poorly polarized lytic granule exocytosis. F-actin stabilization by jasplakinolide did not result in an increase in F-actin intensity (Figures 2A and 2B). Treatment with an intermediate concentration of 1 μ M led to a mild decrease in F-actin intensity, possibly through competition of the drug with phalloidin-AF488 for F-actin binding (Bubb et al., 1994). Different from latrunculin B, jasplakinolide appeared to not alter the cell morphological descriptors, implying that F-actin polymerization, but not stabilization, is crucial for NK cell IS assembly. A higher concentration of 2.5 μ M jasplakinolide was tested but discarded because of its apparent detrimental

effect on cell viability. Treatment with jasplakinolide elicited an increase in the perforin-related descriptors, corresponding to more, larger, and brighter granules. In contrast to latrunculin B, jasplakinolide treatment was associated with severe impairment of lytic granule exocytosis (Figure 2C), indicating accumulation of lytic granules at the synapse plane. This confirms that actin turnover is required for lytic granule exocytosis (Carisey et al., 2018; Lyubchenko et al., 2003). Inhibition of acto-myosin contractility via the Rho-associated protein kinase (ROCK)-myosin pathway using blebbistatin or Y-27632 yielded overlapping effects with a mild increase in F-actin intensity and mild decrease in cell roundness (Figures 2A and 2B), suggesting minor control over overall assembly of the IS. Treatment with blebbistatin and Y-27632 increased the number of granules detected at the synapse and the area they occupied but had no effect on LAMP-1 surface exposure (Figure 2C; Figure S2). In partial contrast with previous reports regarding the role of myosin IIA in promoting lytic granule fusion with the cell membrane (Andzelm et al., 2007; Sanborn et al., 2009), our data suggest that acto-myosin contractility contributes to lytic granule positioning at the IS without affecting degranulation per se. Given the prominent role of actin polymerization in controlling the IS, we compared the relative contribution of actin branching and actin filament elongation, employing inhibitors of the ARP2/3 complex (CK-869), WASP (wiskostatin), and formins (SMIFH2). CK-869 treatment led to a concentration-dependent

decrease in F-actin intensity and cell shape alterations, whereas wiskostatin and SMIFH2 treatments resulted in a modest increase in F-actin intensity without distinct alteration of cell shape descriptors (Figures 2A and 2B). A specific property of CK-869 treatment was a reduction in the number of and area covered by perforin granules, possibly reflecting the inability of CK-869-treated cells to polarize lytic granules toward the stimulatory surface because of defective IS assembly. In agreement, this was accompanied by a dose-dependent reduction in lytic granule exocytosis (Figure 2C; Figure S2), as reported previously (Randzavola et al., 2019). Low-concentration wiskostatin and SMIFH2 treatments resulted in an increase in perforin intensity and area, supporting a role of WASP and formins in lytic granule positioning at the IS (Houmadi et al., 2018; Butler and Cooper, 2009). The effect of these treatments on degranulation was not conclusive because cell toxicity was observed over the 4 h of the LAMP-1 surface exposure experiments. Overall, our comparative analysis shows that lytic granule distribution at the NK IS is dependent on the integrity of various facets of actin dynamics, supporting the notion that multiple actin-dependent steps control lytic granule docking and exocytosis (Mace et al., 2014). We further identify the ARP2/3 complex as playing a dominant role in driving actin polymerization, sustaining IS assembly, and controlling lytic granule positioning.

High-definition morphological profiling of NK cells upon drug treatment

To further enrich the morphological analysis of the IS in the context of drug treatment, the initial set of curated descriptors was expanded to include all available measurements. These included various descriptors of cell shape, staining intensity distributions and textures, as well as descriptors of heterogeneity, such as standard deviation of the aforementioned measurements per field of view, because they were found to be informative in other morphological profiling studies (Breinig et al., 2015; Bryce et al., 2019; Caldera et al., 2019; Rohban et al., 2019). From 1,898 initial measurements, a set of 383 morphological descriptors was retained following filtering of non-informative and redundant descriptors. By opposition to the curated morphological descriptors used previously, we further refer to this new set as comprehensive morphological descriptors. To visualize and quantify the significance of morphological changes upon drug treatment compared with untreated cells, we applied a uniform manifold approximation and projection (UMAP) dimensionality reduction to these morphological profiles. This allows visualization of the relation between the morphology displayed on each field of view in a 2D morphological space, summarizing the variation on all the 383 comprehensive morphological descriptors (Figure S3A). We sorted these measurements based on the cellular compartment (nucleus, cytoplasm, or granules) and the concept they described, such as granularity or shape. By examining the relation between them, we observed no strong clustering, suggesting that the different types of measurements acquired and the different compartments studied provided complementary and non-redundant information about the IS changes occurring across treatments (Figure S3B). This also implied that none of these morphological descriptors were repeating technical confounders, such as experimental plate position effect or cell density, because the confounders were more

similar to each other than to any other descriptor. The morphological space revealed that latrunculin B, jasplakinolide, and CK-869 treatments were clustered away from the untreated cells and from one another, most likely because of these drugs having prominent and distinct effects on the ability of NK-92 cells to assemble the IS (Figures 3A, 3B, and 3E). In comparison, morphologies of cells treated with blebbistatin, Y-27632, wiskostatin, and SMIFH2 appeared to be less distinct from the untreated condition and to cluster in close vicinity of one another (Figures 3C, 3D, 3F, and 3G). The three concentrations assayed per treatment fell into distinct sub-clusters, clearly indicating dose-dependent effects, as detailed for CK-869 and SMIFH2 (Figures S3C and S3D). All drug-evoked morphological profiles were found to be significantly distant from the untreated state. Indeed, the median robust Mahalanobis distances between the fields of view of treated cells and their matching negative controls are larger than expected at random (Figure S3E; Hutz et al., 2013; Rousseeuw and Leroy, 1987). To get insight into the nature of the morphological changes identified in this experiment and pinpoint which of these changes are hallmarks of the chemical treatments, we trained a random forest classifier (Breiman, 2001). This tells us (1) whether the changes are strong enough to result in good model performance and (2) which features are essential to delimit each class. The image set was split to carry out parameter optimization and validate the performance of the model by cross-validation. The classifier achieved good performance, as shown on the confusion matrix (Figure 3H), with an F_1 score and an accuracy of 0.89 and 89%, respectively, which highlights the relevance of the comprehensive morphological descriptors to characterize treatment effects. Most drugs were predicted with high accuracy based on the corresponding image measurements, whereas the morphological effects of blebbistatin and Y-27632 were more challenging to distinguish, in line with their highly related mechanism of action. The importance of each descriptor for classification was proxied by the average increase in accuracy obtained by including the given variable in a decision tree. In particular, our analysis shows that CK-869 treatment mostly affected nucleus and cytoplasm shape descriptors (Figure 3I), whereas SMIFH2 treatment altered radial intensity distributions in the cytoplasm (Figure 3J). Only four measurements described intensities in the cytoplasm within the comprehensive descriptors. Interestingly, those few descriptors, on average, increased model accuracy the most, strengthening the necessity, but not sufficiency, of actin intensity measurements to profile the IS. Descriptors pertaining to lytic granules also played a determinant role in reinforcing model accuracy, providing further evidence of a tight regulation of lytic granule distribution at the IS. In comparison, we trained a classifier only on the curated descriptors used previously (Figure S3f). The overall accuracy of 69% and F_1 score of 0.70 confirmed that these descriptors characterize relatively well the morphological effects of the actin drugs, but the unbiased comprehensive approach offers a considerably higher definition of the morphological description of the IS. It also confirmed F-actin intensity as a major discriminating feature and identified cell eccentricity and roundness as key features to account for the morphological alterations induced by the drugs (Figure S3G). Our data therefore demonstrate the ability of the unbiased profiling to identify relevant

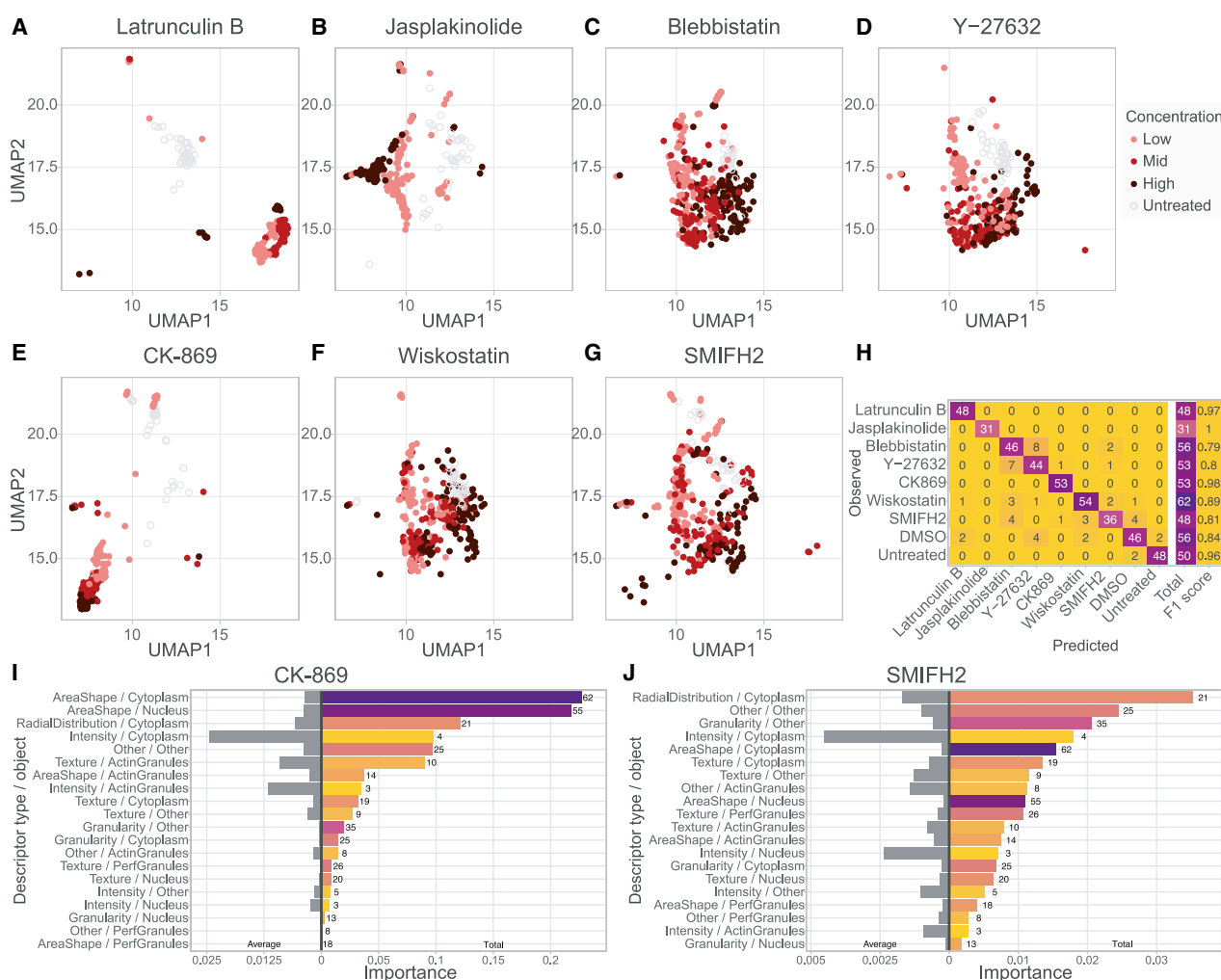


Figure 3. Morphological profiling of the NK cell IS upon drug treatment

(A–G) NK-92 cells were treated with the indicated drugs. Corresponding images were analyzed with CellProfiler for an array of measurements and visualized using UMAP to position drug-treated cells with respect to untreated cells from the same row.

(H) Confusion matrix and class-wise performance on held-out data of a random forest model trained to predict drug treatment based on the morphology of NK-92 cells seeded on ICAM-1, anti-NKp30, and anti-NKp46.

(I and J) Total and average importance for prediction of morphological features per measurement type and biological object described of NK-92 cells seeded on ICAM-1, anti-NKp30, and anti-NKp46 and treated with (I) CK-869 or (J) SMIFH2.

In (H)–(J), a darker color indicates a higher number of images or descriptors for a given box, and the corresponding count is provided alongside.

spatially localized events and characterize perturbed cell states with high-definition power.

Morphological profiling of primary human NK cells upon drug treatment

We next explored the applicability of HCI to study primary human lymphocytes. For that purpose, NK cells were purified from the peripheral blood of three normal donors, treated with four concentrations of CK-869 or SMIFH2, and stimulated with ICAM-1 and anti-NKp30/NKp46 Abs (Figure 4A). We disregarded PLL as a control surface because primary NK cells appeared to strongly remodel their morphology over this surface, in line with recent observations (Santos et al., 2018). The stimulatory

surface promoted an actin-rich IS with lytic granules concentrated in one area on the side of the nucleus. The four concentrations of CK-869 caused a marked decrease in F-actin intensity in the NK cells from the three donors, demonstrating the capacity of morphological profiling to detect actin cytoskeleton alterations in primary lymphocytes. Notably, the area covered by the perforin granules, taken as an absolute value or divided by the cell area, was increased in the CK-869-treated NK cells from the three donors, showing a clear dose-dependent response (Figure 4B). This effect is opposite to what was measured in NK-92 cells, highlighting contrasting responses of model cell lines and primary cells. Moreover, the four concentrations of SMIFH2 also caused a decrease in F-actin intensity in the

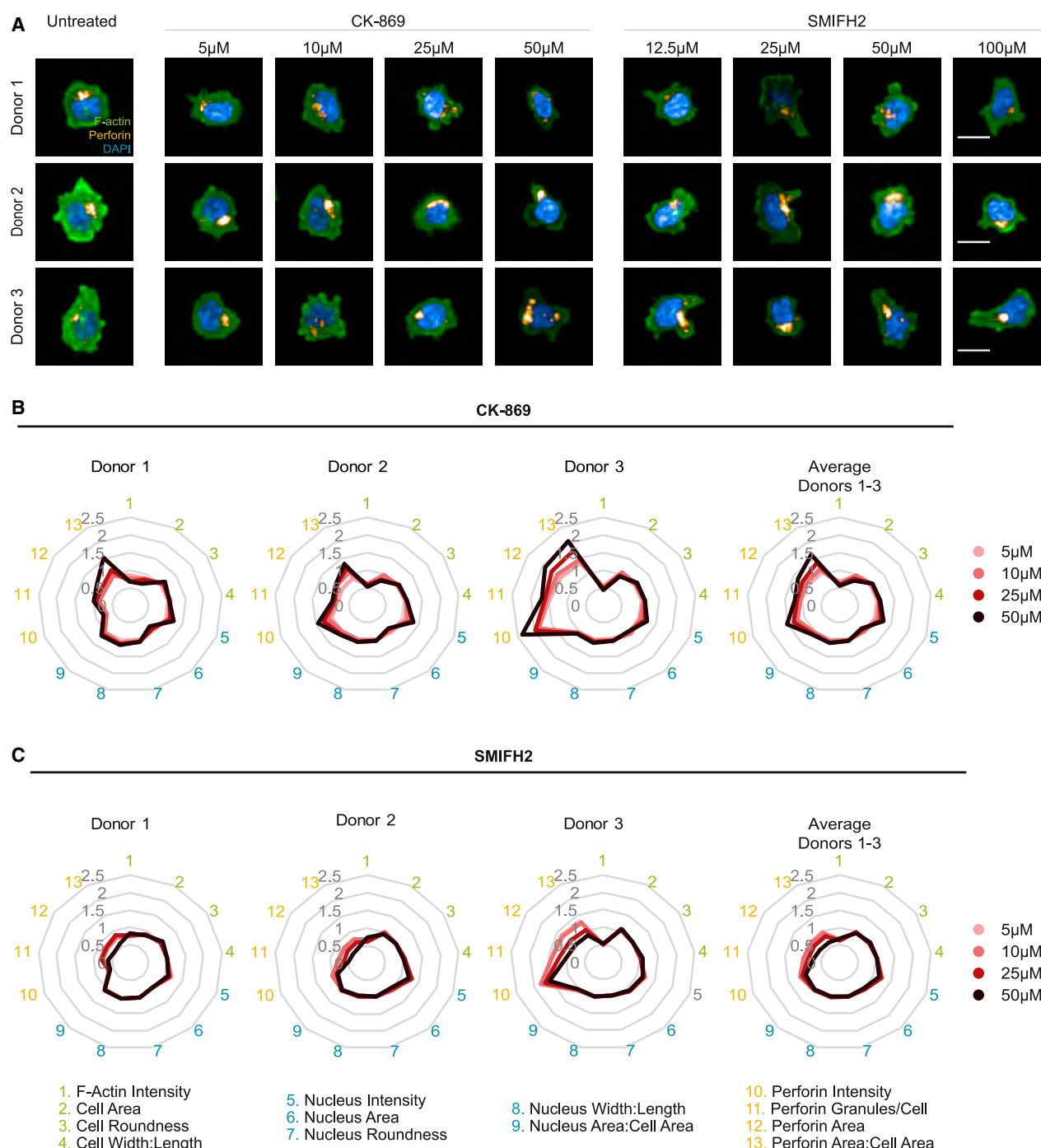


Figure 4. CK-869 and SMIFH2 treatments alter IS architecture and lytic granule polarization in primary NK cells

(A) Representative images of primary NK cells isolated from peripheral blood mononuclear cells (PBMCs) of three normal donors seeded on ICAM-1, anti-NKp30, and anti-NKp46, stained for F-actin (green), perforin granules (yellow), and nuclei (blue) and left untreated or treated with four concentrations of CK-869 or SMIFH2. Scale bars, 10 μ m.

(B and C) Graphs representing the fold changes of IS parameters of primary NK cells treated with (B) CK-869 and (C) SMIFH2 with respect to untreated controls. The data represent the mean of 4 replicates for each drug concentration (60–409 cells).

NK cells from the three donors, highlighting the importance of formins for actin remodeling at the IS of primary NK cells (Figure 4C). SMIFH2 treatment also strongly affected the distribution of perforin granules. Remarkably, dose-dependent reductions in perforin granule number and covered area were observed in NK cells from the first two donors and to a lesser extent in the NK cells from the third donor. Therefore, although inhibition of actin branching and actin elongation yields comparable net decreases in polymerized actin, these treatments differentially affect the distribution of lytic granules. Those observations highlight the potential of HCI to identify complex relationships between cytoskeleton remodeling and organelle trafficking. In addition, HCI reveals features underlying inter-donor variability upon stimulation and treatment of lymphocyte populations.

Distinct IS defects in CD8⁺ T cells from ARPC1B- and WASP-deficient individuals

To broaden the scope of application of HCI and assess its relevance in hematological disease settings, we implemented morphological profiling of CD8⁺ T cells isolated from individuals suffering from two related inborn errors of immunity (IEIs) caused by mutations in *ARPC1B* and *WASP*, which encode a subunit of the ARP2/3 complex and the ARP2/3 activator WASP, respectively. Cells from three ARPC1B-deficient and three WASP-deficient individuals and three normal donors were stimulated with ICAM-1 anti-CD3 Abs and stained for F-actin, perforin, the nucleus, and CD8 (for exclusion of contaminating non-CD8⁺ T cells). Analysis of 13 curated morphological descriptors highlighted that F-actin intensity was decreased in cells from ARPC1B- and WASP-deficient individuals compared with cells from normal donors (Figures 5A and 5B). This decrease was observed in all individuals (Figures S4A–S4D). When examining the mean values of the descriptors pertaining to shape and the nucleus, T cells from ARPC1B-deficient and WASP-deficient individuals did not appear to display striking alterations (Figure 5B; Figures S4A–S4D). Remarkably, the number of detected perforin granules and, consequently, the average fraction of the cell area they occupied was increased in T cells from ARPC1B- and WASP-deficient individuals (Figure 5B). This was the case for all three WASP-deficient individuals and two of three ARPC1B-deficient individuals (Figures S4A–S4D). The effect of these natural deficiencies on lytic granule positioning is reminiscent of the increase in perforin descriptors observed above upon treatment of primary NK cells with the ARP2/3 inhibitor CK-869. Our findings also align with a previous characterization of IS defects in WASP-deficient (Houmadi et al., 2018) and ARPC1B-deficient individuals (Brigida et al., 2018; Randzavola et al., 2019). We took advantage of the side-by-side HCI analysis of T cells from the 2 disease entities to identify specific traits. We selected a comprehensive set of 306 informative descriptors we projected onto a morphological space with UMAP. This analysis evidenced a marked segregation of affected individuals from control donors but also among the 2 disease entities (Figure 5C). This indicates that, beyond reduced F-actin intensity and altered distribution of lytic granules, additional morphological descriptors might be differentially altered in ARPC1B- and WASP-deficient CD8⁺ T cells. Applying a random forest model, we then tested the ability of our approach to distinguish cells from normal donors and

from each of the two deficiencies. Near-perfect classification was obtained on the validation set of images, with an overall accuracy of 96% and F_1 score of 0.96 (Figure 5D), confirming that the ARPC1B and WASP deficiencies were associated with distinct morphological alterations. The categories of descriptors selected most often by the random forest model indicated that ARPC1B-deficient T cells were mostly recognized on the basis of descriptors pertaining to the cell periphery (referred to as “cytoplasm” and corresponding to the cell mask extending beyond the nucleus projection) (Figures 5E and 5F). These descriptors included mean and standard deviation of radial distribution of actin intensities as well as shape- and texture-related parameters. These descriptor classes also accounted partly for the discrimination of WASP-deficient T cells. However, these cells were mostly recognized on the basis of descriptors pertaining to the cell center (referred to as “nucleus” and corresponding to the nucleus projection mask). These descriptors included shape- and texture-related parameters. This ranking of discriminative descriptor classes reveals distinct morphological signatures in T cells with the related ARPC1B and WASP deficiencies that suggest non-overlapping spatial distribution of associated defects in the context of IS assembly.

Altered lytic granule distribution in ARPC1B-deficient CD8⁺ T cells revealed by single-cell analysis

CD8⁺ T cells isolated from an ARPC1B-deficient individual have been shown recently to be unable to kill target cells, and this impairment was associated with a combination of defects in lamellipodium formation, centrosome polarization, lytic granule dispersion, and expression of TCR, CD8, GLUT1, and granzyme B (Randzavola et al., 2019). Here we sought to exploit the confocal resolution and single-cell granularity of our HCI approach to investigate in more detail how defective actin remodeling may relate to aberrant distribution of lytic granules in the context of ARPC1B deficiency. We confirmed that ARPC1B-deficient CD8⁺ T cells have reduced cytotoxic activity (Figure S5D) with evidence of aberrant morphology upon contact with target cells (Figure S5E). However, the CD8⁺ T cells from the ARPC1B-deficient individuals considered in our study expressed normal levels of the TCR, CD8, perforin, and granzyme B (Figure S5C), indicating that the killing defect was not related to aberrant expression of these molecules. We first took advantage of the confocal resolution of our imaging approach to quantify lytic granule distance from the synaptic plane. Eight confocal slices spanning 3.5 μm above the cell-substrate interface were acquired in cells stained for perforin, F-actin, and the nucleus. Our analysis indicates that lytic granules are detected all along the considered Z-planes with enrichment in z layers 3 to 5 corresponding to the layer 1–2 μm from the synapse plane (Figure 6A). Clearly, polarization of the lytic granules toward the synapse plane was not altered in CD8⁺ T cells from the 2 ARPC1B-deficient individuals, implying that ARPC1B deficiency in the 2 studied individuals did not prevent lytic granule positioning toward the IS. Next we exploited the power of our HCI approach in collecting morphological descriptors pertaining to individual cells to extract relations between actin and perforin features at a single-cell level. Using multiple linear regression, we estimated the relationship between key actin-based descriptors and lytic granule

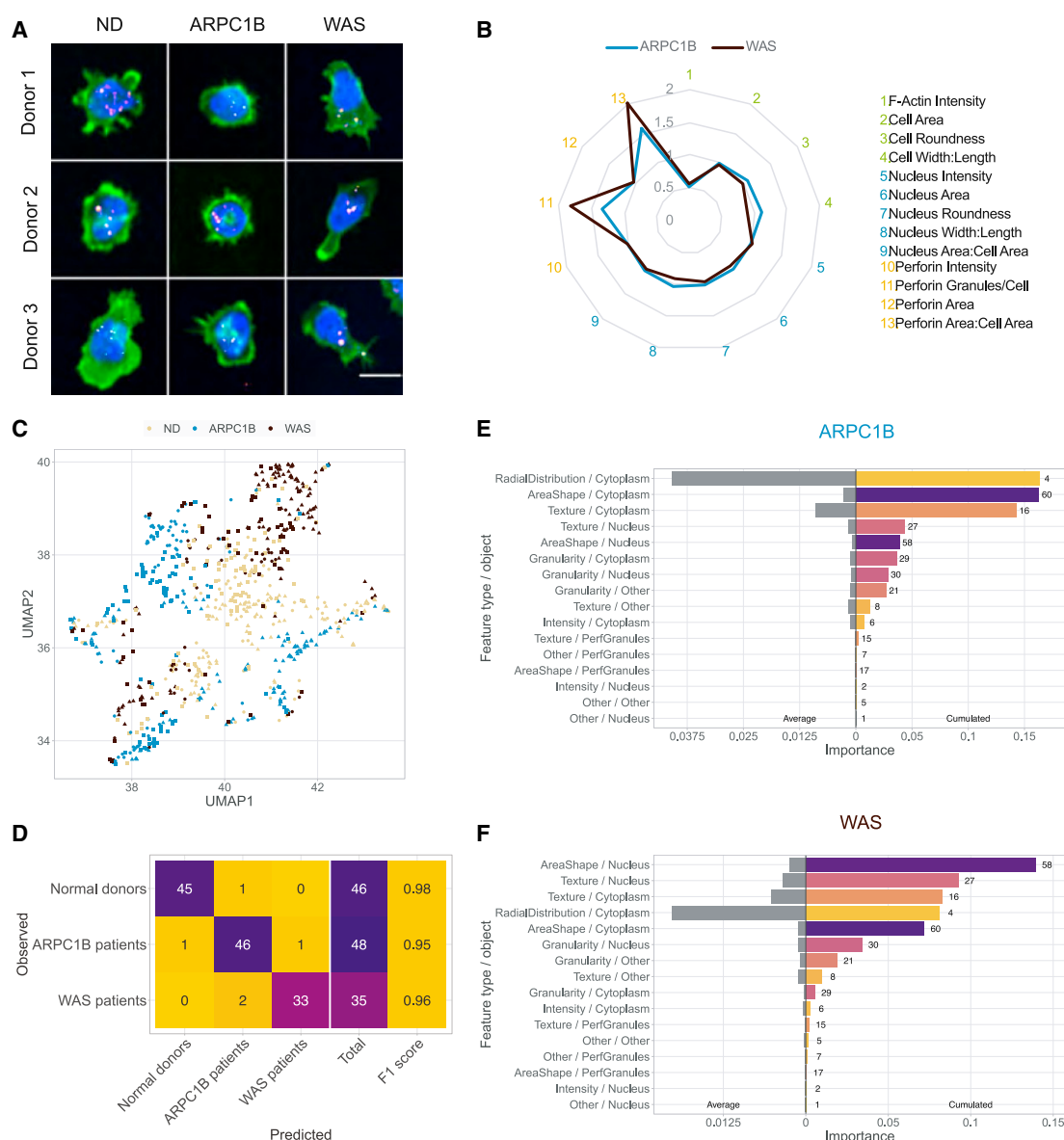


Figure 5. HCI of the IS in WASP- and ARPC1B-deficient CD8⁺ T cells

(A) Representative images of CD8⁺ T cells from normal donors and affected individuals, seeded on ICAM-1/anti-CD3 Abs and stained for F-actin (green), perforin granules (yellow), CD8 (red), and nuclei (blue). Scale bars, 10 μ m.

(B) Average characteristics of the IS of CD8⁺ T cells from affected individuals, represented as fold change with respect to the average of three normal donors. The data represent the mean of eight replicates for each sample (24–443 cells per replicate, 2,890–4,881 cells per condition).

(C) UMAP projection of CD8⁺ T cell morphological profiles. Different shapes denote different samples under each condition (dots, squares, and triangles for the first, second, and third donor/affected individual, respectively).

(D) Confusion matrix and class-wise performance on held-out data of a random forest model trained to discriminate between affected individuals and normal donors.

(E) Total and average importance of morphological features per measurement type and biological object described for prediction of ARPC1B deficiency status.

(F) Total and average importance of morphological features per measurement type and biological object described for prediction of WASP deficiency status.

In (D)–(F), a darker color indicates a higher number of images or descriptors for a given box, and the corresponding count is provided alongside.

positioning in individual normal donor CD8⁺ T cells at the Z-plane the most proximal to the cell-substrate interface (Figure 6B). The mean minimal distance between lytic granules and the cell edge was considered to account for the radial dispersion of lytic gran-

ules. We wanted to find out whether descriptors pertaining to F-actin distribution and cell shape might correlate with lytic granule dispersion. Eight independent variables were found to significantly contribute to the capacity of our model to predict

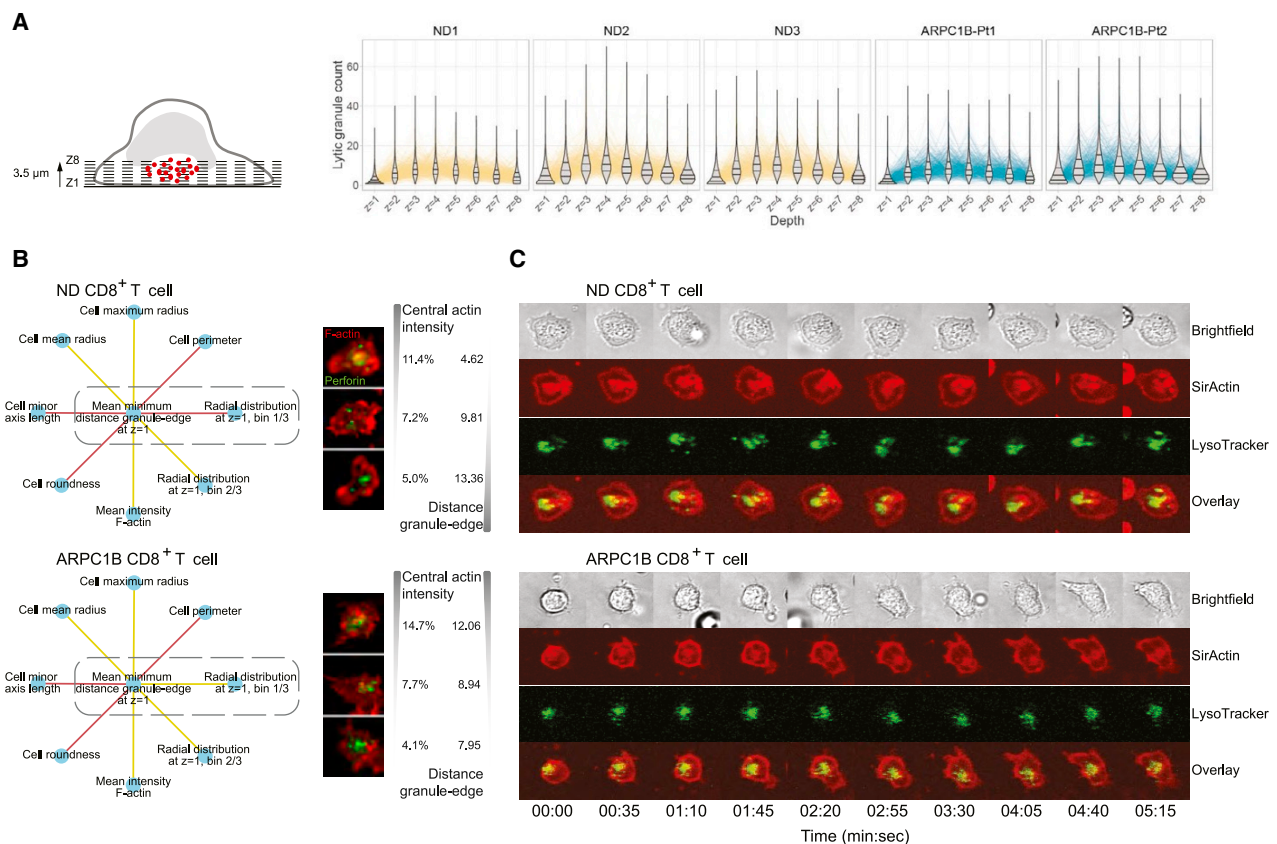


Figure 6. In-depth characterization of ARPC1B deficiency through single-cell analysis of the 3D distribution of lytic granules, granule-actin covariate analysis, and live imaging

(A) Distribution over consecutive focal planes of individual perforin⁺ objects identified as lytic granules in CD8⁺ T cells from 3 healthy donors and 2 ARPC1B-deficient individuals upon interaction with ICAM-1 and anti-CD3 Abs.

(B) Positive (green) or negative (red) associations using linear regression of the average minimum distance between the lytic granules and the cell edge based on eight actin-based morphological descriptors at the synapse plane of the indicated cells. Representative images of healthy donor and ARPC1B-deficient CD8⁺ T cells, showing the relationship between lytic granule and F-actin distribution.

(C) Snapshots of representative video recordings of healthy donor and ARPC1B-deficient CD8⁺ T cells interacting with an ICAM-1 and anti-CD3 Ab-coated surface and stained with silicon rhodamine (SiR)-actin (red) and LysoTracker (green).

lytic granule dispersion in individual cells ($r^2 = 0.2886$; Table S3). In particular, our analysis revealed that the radial distribution of F-actin was associated with lytic granule dispersion even when accounting for geometrical confounders. More precisely, the higher the fraction of F-actin intensity at the synapse center, the higher the dispersion of lytic granule toward the cell edges (Figure 6B). This is in line with the notion that partial depletion of F-actin at the IS center is required for lytic granule exocytosis at this site. When this analysis was applied to ARPC1B-deficient T cells, the association of lytic granule dispersion with 7 of the 8 F-actin/shape descriptors held true (Figure 6B; Table S3). However, the association with F-actin density at the IS center was inverted, meaning that higher central F-actin intensity was associated with more central lytic granules (Figure 6B; Table S4). This analysis of parameter correlation at the single-cell level indicates that ARPC1B deficiency disrupts the association between the radial distribution of F-actin and lytic granule positioning. To resolve the temporal aspect of lytic granule disper-

sion in ARPC1B-deficient CD8⁺ T cells, we complemented our HCI 3D analysis by time-lapse microscopy using identical stimulatory conditions. We employed a spinning disk microscopy setting to focus on the 1- to 2- μ m layer above the IS and acquired images every 35 s. Actin dynamics were visualized with SiR-actin, whereas lytic granules were revealed with LysoTracker. In line with the HCI data, ARPC1B-deficient T cells displayed aberrant IS assembly upon interaction with the stimulatory surface with abnormal cell edge dynamics (Figure 6C; Video S1). In healthy donor and ARPC1B-deficient CD8⁺ T cells, lytic granules gathered around a ring-like actin structure, presumably corresponding to the previously described peri-centrosomal F-actin network (Obino et al., 2016). However, this F-actin ring and the bulk of lytic granules appeared to oscillate more in ARPC1B-deficient CD8⁺ T cells and to occupy a less central position compared with their normal counterparts (Figure 6C; Video S1). Our HCI measurements, complemented by live-recording observations, indicate that ARPC1B deficiency does not affect

the overall polarization of the lytic granule pool but affects its dynamic distribution at the synaptic plane. Our analysis points to the aberrant radial organization of the actin cytoskeleton at the IS as a major factor contributing to this defective distribution.

DISCUSSION

By combining automated cell imaging with computational image analysis pipelines, HCI provides novel opportunities to systematically analyze cellular mechanisms (Breinig et al., 2015; Bryce et al., 2019; Rohban et al., 2017). However, the potential of this approach has not yet been explored for the study of immune cells. Here we tailor an HCI approach for high-definition morphological profiling of various human cytotoxic lymphocyte populations and focus on imaging of the IS as a means to capture the responsiveness and effector potential of these cells. We validate our HCI approach by identifying distinct morphological signatures evoked by a panel of actin-targeting drugs. We were thus able to characterize different defects occurring at the IS and how they converge in altering lytic granule positioning. We further reveal the power of our HCI approach to discriminate individual donors on the basis of immune cell morphological traits. We also exemplify the clinical applicability of this approach by identifying cytotoxic lymphocyte aberrations in individuals with severe congenital IELs.

Conveniently, a simple 2D static approach based on adsorption of stimulatory molecules on the surface microwells was sufficient to robustly stimulate assembly of morphological structures qualifying as ISs. We show that various human lymphocyte populations, including model cell lines, cells freshly isolated from the blood, as well as expanded primary cells, can be stained and imaged with an automated confocal microscope at high resolution in a 384-well format, allowing analysis of several samples, activation conditions, and perturbations in parallel. Computationally, we use robust statistics and work at an image-level resolution, typically gathering a few dozen cells imaged over four Z-planes (along the optical axis) representing the 2- μ m section of the cells most proximal to the stimulatory substrate. Although most morphological profiling studies have been limited to average profiles over wells or replicates (Breinig et al., 2015; Caie et al., 2010; Caldera et al., 2019), a few approaches have defined profiles based on single cells (Caicedo et al., 2018; Lu et al., 2019). To compare samples and treatments, we considered the variability in morphology displayed in each image by including measures of dispersion that have been proven to be beneficial for morphological profiles and could potentially be improved by adding a higher-order joint statistical moment (Rohban et al., 2019). We also exploit individual cell heterogeneity to extract relationships among detected descriptors (e.g., F-actin radial distribution and lytic granule dispersion) at single-cell resolution. From an analytical point of view, we elaborate two complementary methods. First, we focus on a pre-defined set of morphological descriptors based on prior knowledge, including cell and nucleus shape parameters as well as intensities of F-actin, LFA-1, and perforin at the synaptic plane. We show that this method can be applied to relatively low numbers of images and provides meaningful identification of discriminative features when comparing experimental conditions. Second, we implement a high-definition and unbiased morphological profiling pipeline from which novel

relevant parameters can be identified and high-performance classifiers can be trained to discriminate cell states corresponding to different stimulations, drug treatments, or genetic defects.

Beyond the methodological advance provided in this study, we present data relevant to understanding lymphocyte activation from a fundamental and medically relevant perspective. Among the pre-defined set of morphological descriptors we identify an increase in F-actin as a hallmark of T and NK lymphocyte stimulation by combinations of ICAM-1 and Abs directed against CD3 or NK receptors, respectively. This is in line with the previously established role of the actin cytoskeleton in driving the cell spreading behavior supporting IS assembly (Fritzsche et al., 2017; Roy and Burkhardt, 2018). Further investigation of the role of actin cytoskeleton remodeling by treatment of NK cells with a drug array reveals distinct morphological alterations upon targeting actin polymerization, depolymerization, and myosin II. Our data also point to converging morphologies induced by some of the drugs with distinct modes of action, possibly related to a limited number of configurations of the cytoskeleton, as described recently in an adherent neuroblastoma cell line (Bryce et al., 2019; Yin et al., 2014). Strikingly, most tested drugs yield prominent alteration of the distribution of perforin-containing granules, indicating that the different facets of actin cytoskeleton dynamics are important to regulate polarized delivery of lytic granules at the IS (Carisey et al., 2018; Hetrick et al., 2013). Because of the distinct morphological profiles observed for each drug and detection of dose-dependent effects in cell lines and primary cells, we envision that such an approach could be applied for screening immunotherapeutic drugs.

Application of morphological profiling to lymphocyte populations reveals a previously unappreciated level of heterogeneity in cellular morphological traits among individuals. When considering the data pertaining to NK cells freshly isolated from the blood, we cannot rule out that morphological differences arise from distinct activation states of cells from different donors. However, *in vitro* stimulation and expansion of T lymphocytes, which are expected to robustly drive cells toward a differentiated phenotype (Litterman et al., 2014), were also associated with distinct morphological traits. Further analysis of larger cohorts of donors and sorted subpopulations of lymphocytes, which would be compatible with the approach developed here, will be required to precisely evaluate the degree of morphological heterogeneity among individuals and lymphocyte subsets. The sensitivity of the HCI approach is also highlighted by its ability to distinguish morphological alterations in T lymphocyte populations isolated from individuals suffering from related IELs. We show that the limited set of IS morphological descriptors (because they might be selected for a classic analysis) suggests an overlap of IS alterations in CD8⁺ T cells of WASP-deficient and ARPC1B-deficient individuals. However, implementation of high-definition morphological profiling unravels distinct traits for each disease entity, with WASP and ARPC1B deficiencies affecting features of the IS center and periphery, respectively.

Focusing on the recently identified ARPC1B deficiency, we further exemplify the potential of HCI to extract robust quantification of cellular events with high spatial resolution at the single-cell level. In particular, staining of perforin combined with acquisition of 8 confocal planes covering a deep layer of the IS allows

analysis of lytic granule positioning toward the cell-substrate interface. Our 3D analysis indicates that ARPC1B deficiency does not impair lytic granule polarization in the context of 2D synapses. Further analysis of the XY dispersion of lytic granules at the IS plane across normal individual cells reveals numerous correlates with morphological and F-actin-related descriptors, pointing to the covariates that contribute to lytic granule distribution. Interestingly, ARPC1B deficiency appears to reverse the correlation between dispersion of lytic granules and radial distribution of F-actin. These findings motivated complementary investigation of lytic granule and F-actin dynamics by live imaging. Our data indicate that the reduced cytotoxic activity of CD8⁺ T cells from ARPC1B-deficient individuals is mostly related to loss of coordination between F-actin distribution and lytic granule distribution at the IS plane. Distinct secondary events, such as compensatory mechanisms (e.g., the level of ARPC1A expression) or mutations in other modifier genes, might differentially compensate for the ARPC1B-related defects in the different individuals and explain the more restricted defects we observed in our study (Randzavola et al., 2019; Somech et al., 2017). Our focused study of ARPC1B deficiency illustrates the potential of HCI to provide guidance for implementation of complementary low-throughput assays to assess morphological and functional defects. At this stage, we cannot generalize the case of ARPC1B deficiency in establishing a systematic relationship between IS alteration and functional defects. However, it is interesting that multiple IELs have been found by us and others to be associated with IS defects and functional impairment (Gil-Krzewska et al., 2018; Kalinichenko et al., 2021; Mace and Orange, 2014; Pfajfer et al., 2017; Salzer et al., 2016). Previous reports have also shown that IELs where the IS is defective fail to eliminate target cells (Mace and Orange, 2014; Mizesko et al., 2013; Pfajfer et al., 2017). Systematic analysis of multiple such pathologies and corresponding cellular models would certainly provide a unique opportunity to establish rules linking morphology to function.

Here we provide an innovative HCI approach that surpasses current approaches in unbiasedly detecting morphological traits of lymphocyte populations upon activation. We thoroughly report the experimental and computational methods and provide all scripts used in the analysis to maximize the reproducibility of the approach developed here. Further research will be needed to leverage application of HCI to different blood-derived cell subsets, with foreseeable translation into the fields of cancer therapy and personalized medicine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Cell lines and primary cells
 - Patients

● METHODS DETAILS

- Flow cytometry-based NK-92 cell degranulation
- Cytotoxicity assay
- Live microscopy
- Phenotypic analysis
- Western blotting
- Staining for high-content imaging
- Image acquisition and processing

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Data processing and visualization
- Robust morphological perturbation value
- Random forest models
- Regression of morphological descriptors

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.109318>.

ACKNOWLEDGMENTS

The authors thank Fatima-Ezzahra L'Faquhi-Olive, Valérie Duplan-Eche, Anne-Laure Iscache, and Lydia De La Fuente-Vizuet from the cytometry facility of INFINITY as well as Sophie Allart and Simon Lachambre from the cell imaging facility of INFINITY. We also wish to thank Isabelle Fernandes from the Organo platform of the IRSD and Muriel Quaranta-Nicaise, Michael Caldera, Raúl Jiménez Heredia, and Marianne Guisset for discussions and technical advice and Alessandro Aiuti, Marco Gattorno, and Stefano Volpi for cell lines derived from affected individuals. This work was supported by the Vienna Science and Technology Fund (WWTF-LS16-060 to K.B., J.M., and L.D. and VRG15-005 to J.M.), the INSERM Plan Cancer program (C15092BS to L.D.), the Association Laurette Fugain (to L.D.), the CNRS IRP program (project Systact to L.D.), and the Austrian Academy of Science through DOC Fellowship 25365 (to J.H.).

AUTHOR CONTRIBUTIONS

Y.G. performed experiments, analyzed data, and wrote the paper. L.V. designed and performed image data analyses and wrote the paper. A. Kamnev, L.P., J.H., and A. Kalinichenko performed experiments and analyzed data. A.K.M. generated and analyzed cells derived from affected individuals. A.R. contributed to image acquisition. A.F., J.M., and K.B. participated in research design and scientific discussions. L.D. designed the research, supervised the analysis pipeline, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 6, 2020

Revised: February 25, 2021

Accepted: June 7, 2021

Published: July 6, 2021

REFERENCES

- Andzelm, M.M., Chen, X., Krzewska, K., Orange, J.S., and Strominger, J.L. (2007). Myosin IIA is required for cytolytic granule exocytosis in human NK cells. *J. Exp. Med.* 204, 2285–2291.
- Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32.
- Breinig, M., Klein, F.A., Huber, W., and Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.* 11, 846.

- Brigida, I., Zoccolillo, M., Cicalese, M.P., Pfajfer, L., Barzaghi, F., Scala, S., Oleaga-Quintas, C., Álvarez-Álvarez, J.A., Sereni, L., Giannelli, S., et al. (2018). T-cell defects in patients with *ARPC1B* germline mutations account for combined immunodeficiency. *Blood* 132, 2362–2374.
- Brown, A.C.N., Oddos, S., Dobbie, I.M., Alakoskela, J.-M., Parton, R.M., Eissmann, P., Neil, M.A.A., Dunsby, C., French, P.M.W., Davis, I., and Davis, D.M. (2011). Remodelling of cortical actin where lytic granules dock at natural killer cell immune synapses revealed by super-resolution microscopy. *PLoS Biol.* 9, e1001152.
- Bryce, N.S., Failes, T.W., Stehn, J.R., Baker, K., Zahler, S., Arzhaeva, Y., Bischof, L., Lyons, C., Dedova, I., Arndt, G.M., et al. (2019). High-Content Imaging of Unbiased Chemical Perturbations Reveals that the Phenotypic Plasticity of the Actin Cytoskeleton Is Constrained. *Cell Syst.* 9, 496–507.e5.
- Bubb, M.R., Senderowicz, A.M., Sausville, E.A., Duncan, K.L., and Korn, E.D. (1994). Jasplakinolide, a cytotoxic natural product, induces actin polymerization and competitively inhibits the binding of phalloidin to F-actin. *J. Biol. Chem.* 269, 14869–14871.
- Butler, B., and Cooper, J.A. (2009). Distinct Roles for the Actin Nucleators Arp2/3 and hDia1 during NK-Mediated Cytotoxicity. *Curr. Biol.* 19, 1886–1896.
- Cabana, E., Lillo, R.E., and Laniado, H. (2019). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Stat. Papers*, Published online November 20, 2019. <https://doi.org/10.1007/s00362-019-01148-1>.
- Caicedo, J.C., McQuin, C., Goodman, A., Singh, S., and Carpenter, A.E. (2018). Weakly Supervised Learning of Single-Cell Feature Embeddings. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2018, 9309–9318.
- Caie, P.D., Walls, R.E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M.E., and Carragher, N.O. (2010). High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol. Cancer Ther.* 9, 1913–1926.
- Caldera, M., Müller, F., Kaltenbrunner, I., Licciardello, M.P., Lardeau, C.-H., Kubicek, S., and Menche, J. (2019). Mapping the perturbome network of cellular perturbations. *Nat. Commun.* 10, 5140.
- Carisey, A.F., Mace, E.M., Saeed, M.B., Davis, D.M., and Orange, J.S. (2018). Nanoscale Dynamism of Actin Enables Secretory Function in Cytolytic Cells. *Curr. Biol.* 28, 489–502.e9.
- Cheminant, M., Mahlaoui, N., Desconclois, C., Canioni, D., Ysebaert, L., Dupré, L., Vasconcelos, Z., Malphettes, M., Moshous, D., Neven, B., et al. (2019). Lymphoproliferative disease in patients with Wiskott-Aldrich syndrome: Analysis of the French Registry of Primary Immunodeficiencies. *J. Allergy Clin. Immunol.* 143, 2311–2315.e7.
- de Groot, R., Lüthi, J., Lindsay, H., Holtackers, R., and Pelkmans, L. (2018). Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol. Syst. Biol.* 14, e8064.
- De Meester, J., Calvez, R., Valitutti, S., and Dupré, L. (2010). The Wiskott-Aldrich syndrome protein regulates CTL cytotoxicity and is required for efficient killing of B cell lymphoma targets. *J. Leukoc. Biol.* 88, 1031–1040.
- Dieckmann, N.M.G., Frazer, G.L., Asano, Y., Stinchcombe, J.C., and Griffiths, G.M. (2016). The cytotoxic T lymphocyte immune synapse at a glance. *J. Cell Sci.* 129, 2881–2886.
- Dupré, L., Aiuti, A., Trifari, S., Martino, S., Saracco, P., Bordignon, C., and Roncarolo, M.-G. (2002). Wiskott-Aldrich syndrome protein regulates lipid raft dynamics during immunological synapse formation. *Immunity* 17, 157–166.
- Dustin, M.L., and Cooper, J.A. (2000). The immunological synapse and the actin cytoskeleton: molecular hardware for T cell signaling. *Nat. Immunol.* 1, 23–29.
- Dustin, M.L., and Long, E.O. (2010). Cytotoxic immunological synapses. *Immunol. Rev.* 235, 24–34.
- Fritzsche, M., Fernandes, R.A., Chang, V.T., Colin-York, H., Clausen, M.P., Felce, J.H., Galiani, S., Erlenkämper, C., Santos, A.M., Heddeleston, J.M., et al. (2017). Cytoskeletal actin dynamics shape a ramifying actin network underpinning immunological synapse formation. *Sci. Adv.* 3, e1603032.
- Gil-Krzewska, A., Saeed, M.B., Oszmiana, A., Fischer, E.R., Lagrue, K., Gahl, W.A., Introne, W.J., Coligan, J.E., Davis, D.M., and Krzewski, K. (2018). An actin cytoskeletal barrier inhibits lytic granule release from natural killer cells in patients with Chediak-Higashi syndrome. *J. Allergy Clin. Immunol.* 142, 914–927.e6.
- Grakoui, A., Bromley, S.K., Sumen, C., Davis, M.M., Shaw, A.S., Allen, P.M., and Dustin, M.L. (1999). The Immunological Synapse: A Molecular Machine Controlling T Cell Activation. *Science* 285, 221–227.
- Heigwer, F., Scheeder, C., Miersch, T., Schmitt, B., Blass, C., Pour Jamnani, M.V., and Boutros, M. (2018). Time-resolved mapping of genetic interactions to model rewiring of signaling pathways. *eLife* 7, e40174.
- Hetrick, B., Han, M.S., Helgeson, L.A., and Nolen, B.J. (2013). Small molecules CK-666 and CK-869 inhibit actin-related protein 2/3 complex by blocking an activating conformational change. *Chem. Biol.* 20, 701–712.
- Houmadi, R., Guipouy, D., Rey-Barroso, J., Vasconcelos, Z., Cornet, J., Manghi, M., Destainville, N., Valitutti, S., Allart, S., and Dupré, L. (2018). The Wiskott-Aldrich Syndrome Protein Contributes to the Assembly of the LFA-1 Nanocluster Belt at the Lytic Synapse. *Cell Rep.* 22, 979–991.
- Hutz, J.E., Nelson, T., Wu, H., McAllister, G., Moutsatsos, I., Jaeger, S.A., Bandyopadhyay, S., Nigsch, F., Cornett, B., Jenkins, J.L., and Selinger, D.W. (2013). The multidimensional perturbation value: a single metric to measure similarity and activity of treatments in high-throughput multidimensional screens. *J. Biomol. Screen.* 18, 367–377.
- Joshi, P., and Lee, M.-Y. (2015). High Content Imaging (HCI) on Miniaturized Three-Dimensional (3D) Cell Cultures. *Biosensors (Basel)* 5, 768–790.
- Kalinichenko, A., Perinetti Casoni, G., Dupré, L., Trotta, L., Huemer, J., Galgano, D., German, Y., Haladik, B., Pazmandi, J., Thian, M., et al. (2021). RhoG deficiency abrogates cytotoxicity of human lymphocytes and causes hemophagocytic lymphohistiocytosis. *Blood* 137, 2033–2045.
- Krzewski, K., and Coligan, J.E. (2012). Human NK cell lytic granules and regulation of their exocytosis. *Front. Immunol.* 3, 335.
- Litterman, A.J., Zellmer, D.M., LaRue, R.S., Jameson, S.C., and Largaespada, D.A. (2014). Antigen-specific culture of memory-like CD8 T cells for adoptive immunotherapy. *Cancer Immunol. Res.* 2, 839–845.
- Lu, A.X., Kraus, O.Z., Cooper, S., and Moses, A.M. (2019). Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.* 15, e1007348.
- Lyubchenko, T.A., Wurth, G.A., and Zweifach, A. (2003). The actin cytoskeleton and cytotoxic T lymphocytes: evidence for multiple roles that could affect granule exocytosis-dependent target cell killing. *J. Physiol.* 547, 835–847.
- Mace, E.M., and Orange, J.S. (2014). Lytic immune synapse function requires filamentous actin deconstruction by Coronin 1A. *Proc. Natl. Acad. Sci. USA* 111, 6708–6713.
- Mace, E.M., Wu, W.W., Ho, T., Mann, S.S., Hsu, H.-T., and Orange, J.S. (2012). NK cell lytic granules are highly motile at the immunological synapse and require F-actin for post-degranulation persistence. *J. Immunol.* 189, 4870–4880.
- Mace, E.M., Dongre, P., Hsu, H.-T., Sinha, P., James, A.M., Mann, S.S., Forbes, L.R., Watkin, L.B., and Orange, J.S. (2014). Cell biological steps and checkpoints in accessing NK cell cytotoxicity. *Immunol. Cell Biol.* 92, 245–255.
- Mattiazzi Usaj, M., Sahin, N., Friesen, H., Pons, C., Usaj, M., Masinas, M.P.D., Shuteriqi, E., Shkurin, A., Aloy, P., Morris, Q., et al. (2020). Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. *Mol. Syst. Biol.* 16, e9243.
- McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* 3, 861.
- McQuin, C., Goodman, A., Chernyshev, V., Kamensky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 16, e2005970.
- Mizesko, M.C., Banerjee, P.P., Monaco-Shawver, L., Mace, E.M., Bernal, W.E., Sawalle-Belohradsky, J., Belohradsky, B.H., Heinz, V., Freeman, A.F., Sullivan, K.E., et al. (2013). Defective actin accumulation impairs human

- natural killer cell function in patients with dedicator of cytokinesis 8 deficiency. *J. Allergy Clin. Immunol.* **137**, 840–848.
- Monks, C.R.F., Freiberg, B.A., Kupfer, H., Sciaky, N., and Kupfer, A. (1998). Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86.
- Obino, D., Farina, F., Malbec, O., Sáez, P.J., Maurin, M., Gaillard, J., Dingli, F., Loew, D., Gautreau, A., Yuseff, M.-I., et al. (2016). Actin nucleation at the centrosome controls lymphocyte polarity. *Nat. Commun.* **7**, 10969.
- Papakonstantinou, S., and O'Brien, P.J. (2014). High Content Imaging for the Morphometric Diagnosis and Immunophenotypic Prognosis of Canine Lymphomas. *Cytometry B Clin. Cytom.* Published online February 28, 2014. <https://doi.org/10.1002/cytob.21170>.
- Pfajfer, L., Seidel, M.G., Houmadi, R., Rey-Barroso, J., Hirschmugl, T., Salzer, E., Antón, I.M., Urban, C., Schwinger, W., Boztug, K., and Dupré, L. (2017). WIP deficiency severely affects human lymphocyte architecture during migration and synapse assembly. *Blood* **130**, 1949–1953.
- Priestley, R.S., Cheung, J., Murphy, E.J., Ehebauer, M.T., Davis, J.B., and Di Daniel, E. (2019). A novel high-content imaging-based technique for measuring binding of Dickkopf-1 to low-density lipoprotein receptor-related protein 6. *J. Pharmacol. Toxicol. Methods* **95**, 47–55.
- Randzavola, L.O., Strega, K., Juzans, M., Asano, Y., Stinchcombe, J.C., Gauden-Bone, C.M., Seaman, M.N.J., Kuijpers, T.W., and Griffiths, G.M. (2019). Loss of ARPC1B impairs cytotoxic T lymphocyte maintenance and cytolytic activity. *J. Clin. Invest.* **129**, 5600–5614.
- Rohban, M.H., Singh, S., Wu, X., Berthet, J.B., Bray, M.A., Shrestha, Y., Varelas, X., Boehm, J.S., and Carpenter, A.E. (2017). Systematic morphological profiling of human gene and allele function via Cell Painting. *eLife* **6**, e24060.
- Rohban, M.H., Abbasi, H.S., Singh, S., and Carpenter, A.E. (2019). Capturing single-cell heterogeneity via data fusion improves image-based profiling. *Nat. Commun.* **10**, 2082.
- Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection* (John Wiley & Sons).
- Roy, N.H., and Burkhardt, J.K. (2018). The Actin Cytoskeleton: A Mechanical Intermediate for Signal Integration at the Immunological Synapse. *Front. Cell Dev. Biol.* **6**, 116.
- Salzer, E., Cagdas, D., Hons, M., Mace, E.M., Garncarz, W., Petronczki, Ö.Y., Platzer, R., Pfajfer, L., Bilic, I., Ban, S.A., et al. (2016). RASGRP1 deficiency causes immunodeficiency with impaired cytoskeletal dynamics. *Nat. Immunol.* **17**, 1352–1360.
- Sanborn, K.B., Rak, G.D., Maru, S.Y., Demers, K., Difeo, A., Martignetti, J.A., Betts, M.R., Favier, R., Banerjee, P.P., and Orange, J.S. (2009). Myosin IIA associates with NK cell lytic granules to enable their interaction with F-actin and function at the immunological synapse. *J. Immunol.* **182**, 6969–6984.
- Santos, A.M., Ponjavic, A., Fritzsche, M., Fernandes, R.A., de la Serna, J.B., Wilcock, M.J., Schneider, F., Urbančič, I., McColl, J., Anzilotti, C., et al. (2018). Capturing resting T cells: the perils of PLL. *Nat. Immunol.* **19**, 203–205.
- Somech, R., Lev, A., Lee, Y.N., Simon, A.J., Barel, O., Schiby, G., Avivi, C., Barshack, I., Rhodes, M., Yin, J., et al. (2017). Disruption of Thrombocyte and T Lymphocyte Development by a Mutation in *ARPC1B*. *J. Immunol.* **199**, 4036–4045.
- Somersalo, K., Anikeeva, N., Sims, T.N., Thomas, V.K., Strong, R.K., Spies, T., Lebedeva, T., Sykulev, Y., and Dustin, M.L. (2004). Cytotoxic T lymphocytes form an antigen-independent ring junction. *J. Clin. Invest.* **113**, 49–57.
- Vincent, L., and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598.
- Wu, L.S., and Li, J. (2018). High-Content Imaging Phenotypic Screen for Neurogenesis Using Primary Neural Progenitor Cells. *Methods Mol. Biol.* **1787**, 101–113.
- Yin, Z., Sailem, H., Sero, J., Ardy, R., Wong, S.T.C., and Bakal, C. (2014). How cells explore shape space: a quantitative statistical perspective of cellular morphogenesis. *BioEssays* **36**, 1195–1203.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-human CD3 antibody, clone OKT3	eBiosciences	Cat#16-0037-81; RRID:AB_468854
Mouse anti-human NKp30 antibody, clone P30-15	BioLegend	Cat#325202; RRID:AB_756106
Mouse anti-human NKp30 antibody, clone #210847	R&D systems	Cat#MAB18491; RRID:AB_2149445
Mouse anti-human NKp46 antibody, clone 9-E2	BD PharMingen	Cat#557847; RRID:AB_2149297
Mouse anti-human perforin antibody, clone dG9	Biolegend	Cat#308102; RRID:AB_1085316
Alexa Fluor 647 mouse anti-human perforin antibody, clone dG9	BioLegend	Cat#308110; RRID:AB_493254
PE mouse anti-human granzyme B, clone GB11	BD PharMingen	Cat#561142; RRID:AB_10561690
Pacific Blue mouse anti-human CD8 antibody, clone SK1	BioLegend	Cat#344718; RRID:AB_10551438
PE mouse anti-LAMP1 antibody, clone eBioH4A3	eBioscience	Cat#12-1079-42; RRID:AB_10853326
Alexa Fluor 700 mouse anti-human CD56 antibody, clone B159	BD biosciences	Cat#557919; RRID:AB_396940
Mouse anti-human anti-LFA-1 (CD11a) antibody, clone HI-111	BioLegend	Cat#301202; RRID:AB_314140
Alexa Fluor 647 goat anti-mouse IgG1 cross-adsorbed secondary antibody	ThermoFisher Scientific	Cat#A-21240; RRID:AB_2535809
Alexa Fluor 555 goat anti-mouse IgG2b cross-adsorbed secondary antibody	ThermoFisher Scientific	Cat#A-21147; RRID:AB_141778
Mouse anti-GAPDH antibody, clone 6C5	Santa Cruz Biotechnology	Cat#32233; RRID:AB_627679
Mouse anti-HSP90 antibody, clone F-8	Santa Cruz Biotechnology	Cat#13119; RRID:AB_675659
Anti-WASP rabbit polyclonal antibody	Santa Cruz Biotechnology	Cat#sc-8353; RRID:AB_794067
Anti-ARPC1B rabbit polyclonal antibody	Sigma Aldrich	Cat#HPA004832; RRID:AB_1845044
HRP-conjugated anti-rabbit IgG antibody	Cell Signaling	Cat#7074S; RRID:AB_2099233
Biological samples		
Peripheral blood from healthy donors	Etablissement Français du Sang (EFS Occitanie)	N/A
Peripheral blood mononuclear cells from 3 ARPC1B- and 3 WASP-deficient patients (all male subjects)	San Raffaele Hospital, Milan Italy ; Saint Anna Children's Hospital, Vienna, Austria; Institut Universitaire du Cancer de Toulouse - Oncopole, Toulouse, France	Somech et al., 2017 Brigida et al., 2018 Dupré et al., 2002 De Meester et al., 2010 Cheminant et al., 2019
Chemicals, peptides, and recombinant proteins		
Phalloidin-Alexa Fluor 488	ThermoFisher Scientific	Cat#A12379
4',6-diamidino-2-phenylindole, DAPI	ThermoFisher Scientific	Cat#D1306; RRID:AB_2629482
Recombinant human ICAM-1/Fc chimera	R&D systems	Cat#720-IC
Ionomycin	Sigma-Aldrich	Cat#I0634-5MG
Phorbol 12-myristate 13-acetate, PMA	Sigma-Aldrich	Cat#79346-1MG
Latrunculin B	Abcam	Cat#ab144291
Jasplakinolide	Calbiochem	Cat#420127-50UG
Blebbistatin	Sigma-Aldrich	Cat#B0560-1MG
Y-27632	Calbiochem	Cat#68802-1MG
Arp2/3 Complex Inhibitor II, CK-869	MedChemExpress	Cat#HY-16927-10mM/1ml
Wiskostatin	Sigma-Aldrich	Cat#W2270-5MG
Formin FH2 Domain Inhibitor, SMIFH2	Calbiochem	Cat#344092-10MG

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dimethyl sulfoxide, DMSO	Sigma-Aldrich	Cat#41640-1L-M
Phytohaemagglutinin, PHA	ThermoFisher Scientific	Cat#R30852801
Recombinant human IL-2	Peptotech	Cat#200-02
Aphidicolin	Sigma Aldrich	Cat#A0781
CellTrace Violet	ThermoFisher Scientific	Cat#C34571
Fibronectin	Sigma Aldrich	Cat#F1141-1MG
Cell Tracker green CMFDA	ThermoFisher Scientific	Cat#C7025
SiR-actin	Tebu-bio	Cat#SC001
LysoTracker Blue	ThermoFisher Scientific	Cat#L7525
Protease and Phosphatase Inhibitor Cocktail	ThermoFisher Scientific	Cat#78442
Poly-L-Lysine Solution (0.01%)	Merck	Cat#A-005-C
Paraformaldehyde 16%, EM grade	ThermoFisher Scientific	Cat#50-980-489

Critical commercial assays

EasySep Human CD8 ⁺ T cell enrichment kit	StemCell	Cat#19053
MagniSort Human NK enrichment kit	Invitrogen	Cat#8804-6819
Super Signal West Pico Chemiluminescence Substrate	ThermoFisher Scientific	Cat#34080

Deposited data

Morphological measurements	This manuscript	FigShare: https://doi.org/10.6084/m9.figshare.11619960
----------------------------	-----------------	---

Experimental models: Cell lines

NK-92 lymphoblastic natural killer cell line, originating from the peripheral blood of a male subject	ATCC	Cat#CRL-2407; RRID:CVCL_2142
Jurkat lymphoblastic T cell line, originating from the peripheral blood of a male subject	ATCC	Cat#TIB-152; RRID:CVCL_0367
P815 mastocytoma cell line of murine origin (DBA/2 strain)	ATCC	Cat#TIB-64; RRID:CVCL_2154

Software and algorithms

FlowJo	BD	FlowJo 10.7.2
R	The R Foundation	R 3.5.1
Excel	Microsoft	1902
umap-learn	McInnes et al., 2018	0.3.6
robustbase	CRAN	0.93
randomforest	CRAN	4.6
CellProfiler	McQuin et al., 2018	3.0
ggplot2	CRAN	3.1.1
Harmony high-content imaging and analysis software	PerkinElmer	HH17000012
Analysis notebooks	This manuscript	Zenodo: https://doi.org/10.5281/zenodo.4562363

Other

CellCarrier-384 Ultra Microplates	PerkinElmer	Cat#6057300
-----------------------------------	-------------	-------------

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Loïc Dupré (loic.dupre@inserm.fr).

Materials availability

Raw images of individual experimental sets can be obtained upon request.

Data and code availability

All the CellProfiler pipelines and morphological measurements used in this analysis are made available on FigShare with the DOI 10.6084/m9.figshare.11619960. The analyses can be found and reproduced using the Docker image and scripts provided via Zenodo: <https://doi.org/10.5281/zenodo.4562363>

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and primary cells

Jurkat cells were cultured in RPMI (GIBCO) supplemented with 10% FBS, 100 U/ml penicillin/streptomycin, 1 mM sodium pyruvate, 1X non-essential amino acids and 10 mM HEPES (all from ThermoFisher Scientific). NK-92 cells were cultured according to the recommendations from ATCC. Primary NK cells were purified from freshly isolated PBMCs using the MagniSort Human NK enrichment kit (Invitrogen) and maintained in RPMI supplemented with 5% human serum, 100 U/ml penicillin/streptomycin, 1 mM sodium pyruvate, 1X non-essential amino acids and 10 mM HEPES. Primary CD8⁺ T cells were purified from frozen PBMCs of three healthy donors, three ARPC1B-deficient patients and three WASP-deficient patients by negative selection using the EasySep Human CD8⁺ T cell enrichment kit (StemCell Technologies, Inc.). CD8⁺ T cells were stimulated in RPMI supplemented with 5% human serum, 100 U/ml penicillin/streptomycin, 1 mM sodium pyruvate, 1X non-essential amino acids, 10 mM HEPES, 1 μ g/ml PHA and 100 IU/ml IL-2. CD8⁺ T cells were expanded for further rounds every two weeks with a mixture of irradiated PBMCs from three normal donors. Peripheral blood from healthy donors and patients was obtained in accordance with the 1964 Helsinki declaration and its later amendments or ethical standards. Informed consents were approved by the relevant local Institutional Ethical Committees.

Patients

ARPC1B Patient 1 (ARPC1B-Pt1, male, 15 years) corresponds to Pt1 in the study by Brigida et al. (2018). The patient carries a c.64+1G > C splice donor variant in the *ARPC1B* gene resulting in insertion of 21 nucleotides from intron 2 and usage of an alternative splice site with partial intron retention and maintenance of the reading frame. ARPC1B-Pt1 was initially described as having a subpopulation (21%) of T cells expressing normal levels of ARPC1B (as detected by flow cytometry), possibly resulting from a secondary mutation. However, expanded CD8⁺ T cells failed to express ARPC1B (Figure S5B). ARPC1B-Pt2 (male, 2 months) corresponds to Pt2 in the study by Somech et al. (2017) and to Pt7 in the study by Brigida et al. (2018). ARPC1B-Pt2 carries a 2-bp deletion (n.c.G623DEL-TC) causing a putative frameshift resulting in premature termination (p.V208VfsX20). The ARPC1B protein was not detected (short fragment could theoretically be expressed but remain undetected; see Figure S5B). ARPC1B Pt3 (male, 4 years) is a patient with early disease onset. He carries an homozygous 2-bp duplication (c.613_614dup), leading to a frameshift and premature termination (p.H206YfsX16). ARPC1B protein failed to be detected by western blotting in the T cells from the patient (Figure S5B). WAS-Pt1 (male, 1 year) corresponds to patient WAS1 in the study by De Meester et al. (2010). WAS-Pt1 carries a *WAS* gene nonstop mutation (c.1509A > C), resulting in undetectable WASP expression in the PBMC-derived CD8⁺ T cells (Figure S5A). WAS-Pt2 (male, 5 years) corresponds to patient WAS2 in the study by Dupré et al. (2002). WAS-Pt2 carries a 2-nucleotide deletion (ag) in exon 4 (position 484 to 485) resulting in a stop codon (codon 167). Accordingly, no WASP expression was detected in the expanded CD8⁺ T cells from the patient (Figure S5A). WAS-Pt3 (male, 51 years) corresponds to patient P10 in the study by Cheminant et al. (2019). WAS-Pt3 carries a c.1453G > A missplicing mutation. This patient was reported to express revertant WASP in a minor fraction of peripheral lymphocytes. However we failed to detect WASP expression in the expanded CD8⁺ T cells from the patient (Figure S5A).

METHODS DETAILS

Flow cytometry-based NK-92 cell degranulation

NK-92 cells were incubated for 1 hr with Blebbistatin (5, 10 and 50 μ M), CK-869 (10, 25 and 50 μ M), Jasplakinolide (0.1, 1 and 2.5 μ M), Latrunculin B (0.1, 0.25 and 0.5 μ M), SMIFH2 (50, 100 and 250 μ M), Wiskostatin (10, 50 and 100 μ M), Y-27632 (5, 10 and 25 μ M) and DMSO as control (all drugs from Merck). For stimulation cells were transferred into a 96 well plate (Nunc MaxiSorp) coated overnight at 4°C with 5 μ g/ml anti-NKp30 Ab and 0.4 μ g/ml recombinant human ICAM-1-Fc chimera (R&D Systems) in PBS, briefly spun down and further incubated in the presence of chemical compounds and anti-CD107a Ab for 4 hr. Following incubation, NK-92 cells were stained with anti-CD56 and anti-CD107a Ab for 30 min, washed, fixed and analyzed using LSR-Fortessa flow cytometer.

Cytotoxicity assay

Target P815 cells were stained for 30 min with Cell Tracker green CMFDA (ThermoFisher Scientific) and coated with 0 to 10 μ g/ml anti-CD3 Ab (OKT3, eBiosciences) for one hour at 37°C. They were also treated with 0.2 μ g/ml aphidicolin to prevent proliferation during assessment of killing. CD8⁺ T cells were stained in parallel with CellTrace Violet (CTV, ThermoFisher Scientific). Stained P815 and CD8⁺ T cells were seeded at a 1 to 1 ratio in a 384-well plate precoated with 1 μ g/ml fibronectin and incubated at 37°C, 5% CO₂ for 24 hr. Cells were imaged on an automated spinning disk confocal HCS device (Opera Phenix, PerkinElmer) with a 20x objective, necessitating 25 fields of view per well for full surface coverage. Residual alive target cells were counted automatically with the Harmony software on the basis of their size and CMFDA staining. Percentage of killing was calculated as 100 - (100 x residual alive target cells / unexposed target cells).

Live microscopy

Eight-well chamber slides (Ibidi, ibidi) were coated overnight with 4 $\mu\text{g}/\text{ml}$ recombinant human ICAM-1-Fc chimera (R&D Systems) and 10 $\mu\text{g}/\text{ml}$ anti-CD3 Ab (OKT3, eBiosciences) at 4°C. Cells were stained with 2 μM SiR-actin (tebu-bio) and 1 μM LysoTracker Blue (ThermoFisher Scientific) for 1 hr at 37°C before transfer to the pre-warmed coated slides. Cells were imaged on a Nikon inverted spinning disk confocal microscope equipped with an oil immersion 40X objective (NA 1.3), a sCMOS camera (6.5 $\mu\text{m}/\text{pixels}$) and a temperature and CO₂ controlled chamber. Acquisition of SiR-actin, LysoTracker Blue and DIC were done over multiple wells at 1 image every 35 s for a total of 61 frames.

Phenotypic analysis

Expanded CD8⁺ T cells from normal donors and ARPC1B-deficient patients were stained with fluorochrome-coupled Ab recognizing the extracellular markers CD8 (BioLegend) and LFA-1 (BioLegend) for 30 min at 4°C. Intracellular staining was performed following fixation and permeabilization, with Ab specific for perforin (BioLegend) and granzyme B (BD Pharmingen) for 45 min at 4°C. The data were acquired on a MacsQuant Q10 (Miltenyi) and analyzed with FlowJo. Student's t test was used to calculate significance.

Western blotting

Cell extracts were prepared from 10 × 10⁶ expanded CD8⁺ T cells. Briefly, T cells were washed in PBS and resuspended in RIPA lysis buffer (Millipore), supplemented with protease and phosphatase inhibitor cocktail (ThermoFisher Scientific). After 20 min on ice, lysates were cleared by centrifugation. Laemmli sample buffer (BioRad) was added to supernatants and, after boiling, samples were subjected to SDS-PAGE. Proteins were transferred to nitrocellulose membranes (GE Healthcare). WASP and ARPC1B were detected with, respectively, anti-WASP Ab (H-250, Santa Cruz Biotechnology) and anti-ARPC1B Ab (HPA004832, Sigma Aldrich), followed by secondary HRP-conjugated anti-rabbit Ab (Cell Signaling). Control of protein loading was performed by hybridizing the same membranes with anti-GAPDH or anti-HSP90 Abs (both from Santa Cruz Biotechnology). Detection was performed using the Super Signal West Pico Chemiluminescence Substrate (ThermoFisher Scientific).

Staining for high-content imaging

CellCarrier-384 Ultra tissue culture treated plates (PerkinElmer) were coated with either 0.1 mg/ml poly-L-lysine (Merck) or a combination of 2 $\mu\text{g}/\text{ml}$ recombinant human ICAM-1-Fc chimera (R&D Systems), 1 $\mu\text{g}/\text{ml}$ NKp30 (MAB18491, R&D systems) and 1 $\mu\text{g}/\text{ml}$ NKp46 (557487, BD Biosciences). NK-92 cells were cultured in IL-2 free medium overnight. NK-92 cells or primary NK cells were seeded at 15000 and 5000 cells per well, respectively and incubated for 30 min at 37°C. Cells were fixed with 3% paraformaldehyde (ThermoFisher Scientific) and stained with anti-perforin Ab (dG9, Biolegend) and phalloidin-AF 488 (ThermoFisher Scientific). AF 555-conjugated goat anti-mouse IgG2b Ab (ThermoFisher Scientific) was used to reveal perforin staining. Nuclei were stained with DAPI (ThermoFisher Scientific).

NK-92 cells were treated with 5, 10 and 50 μM Blebbistatin, 10, 25 and 50 μM CK-869, 0.1, 1 and 2.5 μM Jasplakinolide, 0.1, 0.25 and 0.5 μM Latrunculin B, 50, 100 and 250 μM SMIFH2, 10 50 and 100 μM Wiskostatin and 5, 10 and 25 μM Y-27632 (all drugs from Merck) for 30 min at 37°C and washed twice in PBS before seeding onto the plates and letting them adhere for 30 min. The same procedure was applied to primary NK cells treated with 5, 10, 25 and 50 μM CK-869 and 25 50, 100 and 250 μM SMIFH2.

CellCarrier-384 Ultra tissue culture treated plates were coated with either 0.1 mg/ml poly-L-lysine or a combination of 2 $\mu\text{g}/\text{ml}$ recombinant human ICAM-1-Fc chimera (R&D Systems) and 10 $\mu\text{g}/\text{ml}$ anti-CD3 Ab (OKT3, eBioscience). Jurkat cells or untransformed CD8⁺ T cells were seeded at 10000 and 5000 cells per well, respectively and incubated for 15 min at 37°C. Cells were fixed with 3% paraformaldehyde and stained with anti-LFA-1 Ab (clone HI-111, BioLegend) and phalloidin-AF 488 (ThermoFisher Scientific) in permeabilization buffer (eBioscience). AF 647-conjugated goat anti-mouse IgG1 Ab (ThermoFisher Scientific) was used to reveal LFA-1 staining. CD8⁺ T cells were in addition stained with anti-perforin Ab and AF 555-conjugated goat anti-mouse Ab (ThermoFisher Scientific) was used to reveal perforin staining. Where indicated, cells were stained with anti-CD8 Ab instead of anti-LFA-1 Ab. Nuclei were stained with DAPI. Stained cells were kept in PBS at 4°C until imaging. Comparisons of samples or treatments were done within the same staining/acquisition batches, as well as within the same plates to avoid possible bias due to staining intensity variability. To account for possible well-to-well variability, each well was replicated at least 3 times.

Image acquisition and processing

Images were acquired on an automated spinning disk confocal HCS device (Opera Phenix, PerkinElmer) equipped with a 40x 1.1 NA Plan Apochromat water immersion objective and 4 sCMOS cameras (16 bits, 2160 × 2160 pixels, 6.5 μm pixel size), allowing simultaneous acquisition in 4 channels. For each well, 40 automatically selected fields and 8 Z-planes per field (0.5 μm step, starting from the cell-substrate contact plane) were acquired, making use of a spinning microlens disk. For the experiment on primary NK cells, only 31 fields of view and 6 Z-planes were imaged, while 13 fields and 13 Z-planes were imaged in the experiment on WASP- and ARPC1B-deficient patients. Where indicated, stacks of images were combined with maximum projection of 4 focal slices in proximity of the cell-substrate contact plane (z from 2 to 5 with a 0.5 μm step). Image datasets were processed with CellProfiler 3.0 (McQuin et al., 2018) (see [Data and code availability](#)). In brief, the image quality was assessed, the intensities for experiments with high background noise were log-transformed, the illumination on each image was corrected based on background intensities, and DNA precipitations were avoided by multiplying intensities on DAPI channel by phalloidin intensities before segmenting cell

nuclei using global minimum cross entropy thresholding. A secondary segmentation of the cytoplasm was performed using the watershed method (Vincent and Soille, 1991) and global minimum cross entropy thresholding on the phalloidin channel. Image sets with low maximal DNA intensity or showing no nucleus were discarded. Cells having more than 30% of their cytoplasm surface at less than 5 pixels of another cell were removed, in order to ignore clusters of cells and to focus on single cells displaying an IS. Small actin speckles in the cytoplasm at more than 3 pixels from the membrane as well as speckles of perforin and secondary objects spanned around the nuclei by LFA-1 staining were segmented. Additionally, primary NK and expanded CD8⁺ T cells with low CD8 intensities (if stained) or associated with less than two perforin granules were excluded from the analysis. Finally, measurements were acquired for colocalization of these objects, intensities in the nuclei and cytoplasm, granularity on all channels, textural and shape features, intensity distributions, distance and overlap between objects, number of speckles and neighbors less than 10 pixels away. The average and the standard deviation per field of view of these features were kept. This led to 1898 and 2076 morphological features in NK-92 and Jurkat cells respectively. For primary NK cells and expanded patient CD8⁺ T cells, features related to actin speckles were excluded, as they were not found to be informative, resulting in 2386 and 1780 features, respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing and visualization

Analyses in R 3.5.1 with the data visualization package ggplot2 3.1.1 and Microsoft Excel (Version 1902) were subsequently conducted. A smaller set of informative morphological features was selected and the quality of processed images was checked by (i) removing wells with low maximal DNA intensity and cell count, (ii) removing features and images generating missing values and (iii) removing constant features in the study dataset or the subset of negative controls used as reference. From these images passing our quality checks, up to 16 raw summary variables were extracted, based on their interpretability and on their known relevance to describe the IS. The fold changes compared to unstimulated or untreated controls were further reported and displayed in the form of radar charts. On the other hand, for all features, per-image values X were transformed successively with the following functions f_1 and f_2 , with $X_{Control}$ the negative controls in X on which the dataset is normalized:

$$f_1(X) = \log(X + 1 - (X))$$

$$f_2(X) = \frac{X - \text{median}(X_{Control})}{\text{mad}(X_{Control})}$$

To remove redundancy in the set of features used for downstream analyses, we ensured that the selected variables were not excessively linearly correlated. To do so, all features were ordered from highest to lowest median absolute deviation (hence by variation in the experiment compared to negative controls). Starting from the top of this list, all other features linearly correlated to the first feature with a Pearson's coefficient higher than 0.6 were excluded. We sequentially went on with the next remaining feature in the list and iterated until the acquisition of a small and informative set of uncorrelated features. This led to 383 of such features for the drug screen on NK-92 (Figure 3) and 306 for the experiment on ARPC1B- and WASP-deficient patients (Figure 5), coined comprehensive morphological descriptors. This set of features was used for visualization and quantification of the overall morphological changes induced by perturbations. The dimensionality of the data was reduced using the UMAP algorithm (McInnes et al., 2018) to two dimensions for visualizations and three dimensions for computation of the statistical significance of morphological effects in the drug screen on NK-92. This pipeline succeeded in selecting a wide range of features that were not excessively biased by confounders (Figure S3B).

Robust morphological perturbation value

To quantify the significance of overall changes in morphology between a perturbed state and a reference state (healthy or untreated cells), we defined the Robust Morphological Perturbation Value (RMPV). This extends the concept of Multidimensional Perturbation Value (Hutz et al., 2013) which defines a single value summarizing the statistical significance of morphological changes in multidimensional spaces, by using robust statistics and the minimum covariance determinant (Rousseeuw and Leroy, 1987) decreasing the sensitivity to technical (unfiltered artifacts) and biological outliers (images displaying extreme morphologies or uncommon cell states). In brief, the RMPV is obtained for X the set of all filtered and uncorrelated features and X_{WT} the subset of the data corresponding to images of the reference population in five steps. First, the minimum covariance determinant estimator $M(X_{WT})$ is calculated to describe the variation of morphologies observed in the reference set, using its implementation in the R package *robustBase* version 0.93. Second, this value is used to determine R , the robust Mahalanobis distance of each images of X to X_{WT} (Cabana et al., 2019). Third, the median value $\tilde{R} = \text{median}(R)$ was obtained for each drug tested. Fourth, for 2000 iterations the labels of the condition and the reference were randomly permuted to obtain an empirical distribution of \tilde{R} under the assumption that there was no difference between the multivariate location and scatter of the morphological parameters of the perturbation and the reference. Finally, the RMPV is defined as the empirical p value obtained from these distributions after FDR adjustment for testing changes in multiple conditions and indicates the probability of observing at least half of the images displaying morphological changes of a similar intensity if the perturbation was similar to the reference.

Random forest models

We used the random forest algorithm (Breiman, 2001) with its implementation in the R package *randomForest* version 4.6 and built classifier models to characterize the strength and types of the morphological changes observed and identify what was specific of each cellular condition (chemical perturbation or disease state). In brief, the algorithm generates a set of decision trees. Each tree is constructed to minimize the number of misclassified entries of a bootstrap sample of the training data by choosing between a fixed number *mtry* of randomly selected input features at each split. The final prediction of the model is the most common one out of the predictions of all individual trees. We trained the model on the comprehensive set of informative and uncorrelated morphological features measured per image – previously used for dimensionality reduction. Each forest included 1000 decision trees. The dataset was split in 6 folds of equal size (459 feature vectors per fold for the treated NK cells and 126 for the ARPC1B- and WASP-deficient patient cells), each containing all possible classification labels. To select the optimal number *mtry* of variables selected at each split, we incremented the parameter value from 20 to 90 by steps of 10 and assessed the performance using the macro F_1 score as defined below in a 5-fold cross-validation scheme. One extra fold was used as validation set to estimate the performance of the model after selection of the optimal parameters and retraining on all of the 5 folds used for cross-validation. In the case of the drug screen on the NK-92 cell line, we used a similar approach using the 13 features of known relevance in describing the IS as input, and testing *mtry* values from 1 to 13 with steps of 3. Overall, the performance was evaluated using the macro F_1 score:

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i}$$

where n is the number of categories in the classification, and TP_i , FP_i and FN_i are respectively the number of true positives, false positives and false negatives for category i in the validation set. To interpret the feature importance in the prediction, the mean decrease in accuracy obtained when including each feature was extracted, either for the prediction of a given class or overall using micro averaging. The total and average importance of features split in distinct groups based on the type of measurements and biological object described were calculated as well. These feature groups were defined based on the corresponding CellProfiler measurement types and biological objects. Features that did not describe the cytoplasm, nucleus, perforin granules or actin granules were counted in the “Other” biological object category. Similarly, features that did not correspond to the “Texture,” “AreaShape,” “RadialDistribution,” “Granularity” or “Intensity” measurements were grouped under the term “Other.”

Regression of morphological descriptors

For the single-cell morphological measurements of ARPC1B-deficient patients and normal donors (Figures 6A and 6B), we modeled the average radial position of lytic granules per cell, defined as the minimum number of pixels between a perforin granule and the edge of the segmented cell, in terms of eight other measurements based on the actin staining, using a multiple least-square regression with intercept. These covariates were describing the geometry of the cell (“AreaShape_Perimeter,” “AreaShape_FormFactor,” “AreaShape_MeanRadius,” “AreaShape_MaximumRadius,” “AreaShape_MinorAxisLength”), the total average actin intensity (“Intensity_MeanIntensity_CorrActin”) as well as the radial distribution of the actin at the synaptic plane (“RadialDistribution_FracAtD_CorrActin1_1of3” and “RadialDistribution_FracAtD_CorrActin1_2of3”). To note, this corresponds to the intensity fraction when separating the cell in three concentric regions centered on the nucleus and numbered from 1, for the space directly surrounding the nucleus, to 3, for the edge of the cell. The third bin was not included in the model as it is purely redundant to the information contained in bins 1 and 2. Three separate models were obtained across cells from all three normal donors, ARPC1B-Pt1 and ARPC1B-Pt2, respectively (Table S3).

4

Discussion

Taken together, the two articles reproduced above demonstrate how HCI is able to quantify subtle phenotypic changes, which provide valuable information about cellular response to both intrinsic and extrinsic perturbations. Morphological profiling, relying on the quantification of changes in morphological descriptors combined with a careful statistical analysis of such values, takes full advantage of the rich data provided by HCI. The resulting profiles can first be analysed on their own, for instance to study similarity in drug effects as demonstrated in the first paper, or to characterize different actin defects as performed in the second paper. They can also be seen as a global depiction of cellular state as assessed from cellular morphology. Thus, they constitute a layer of valuable information analogous but orthogonal to well-developed OMICS modalities, and can be complemented and integrated with other sources of data. In our analyses, we explored the possibilities of data integration when comparing morphological and PPI network distances, and when contextualizing patient morphological profiles with their molecular phenotyping. Integrative methods could also be developed further in follow-up studies. In this chapter, we broadly discuss the implications, limitations, and avenues opened by the work described in this thesis.

First, we present the developments we envision for the Julia package we implemented. Second, we examine more broadly how profiling analyses may benefit the characterization of immune responses. Third, we briefly describe new HCI projects running in our group that were enabled by the work compiled in the Results chapter, involving both CRISPR knockouts alone or in combination with chemical perturbations. Fourth, we wonder what role morphological profiling will play in the future. We then move on more specifically to an understudied aspect of the profiling workflow, and detail the problems faced when analysing the dispersion of heterogeneous morphological profiles. Finally, we conclude with a section dedicated to the emerging overlap between morphological profiling and network medicine.

4.1 Future of BioProfiling.jl

Initially, the development of BioProfiling.jl was prompted by our own needs. Our research group previously contributed to several studies deriving quantitative metrics from cell imaging, mostly by using CellProfiler for feature extraction and custom scripts implemented in Python for the subsequent data analysis (Caldera et al., 2019; Marini Thian et al., 2020; Salzer et al., 2020). Building on this expertise, I have written and compiled a considerable amount of code, mainly in R and Python, to perform morphological profiling analyses, notably when working on the study presented in section 3.2. By switching from an unstructured collection of code to a dedicated tool following the best practice principles of research software engineering, we achieved several milestones:

- We greatly increased the internal and external re-usability of the code by making it more general and adaptable, writing an extensive documentation, and providing detailed and reproducible examples.
- We made the software more accessible and easier to install, from the main Julia package registry or through Docker images.
- We formalized and streamlined our approach to morphological profile curation into a unified framework, where the user iteratively filters entries and selects features of interest, then applies transformations to the curated dataset. This makes the analysis considerably clearer and reduces the need for error-prone convoluted custom code.
- We improved the tool's performance and features, in part by porting the functions to Julia, allowing parallel computing for the most time-consuming step, and optimizing data transformation methods.
- All methods got extensively validated via unit testing, thus reducing the likelihood of bugs.
- The code became more maintainable and future-proof, thanks to continuous integration, an engineering principle promoting the streamlined merging and testing of code contributions, and to the compatibility features implemented in the central Julia package repository.
- We facilitate the reproducibility of analyses performed with our code, with tagged releases and DOI identifiers.
- Finally, we encourage collaboration by making the project open source and including contribution guidelines and a code of conduct.

On the one hand, the release of our tool and the subsequent publication helped to make the contribution and expertise of our group in the field of bioimage analysis more visible. This already got us involved in more collaborative image-driven projects. On the other hand, the

highest scientific impact of BioProfiling.jl will be obtained if the package gets widely adopted by the community at large. To this end, one of the biggest remaining challenges is to expand our user base outside of our own group and direct collaborators. This would have the additional benefit of providing us with valuable constructive feedback from users to further enhance BioProfiling.jl. Therefore, we have been continuously advertising BioProfiling.jl at venues dedicated to Julia packages, such as the JuliaCon 2021, and to bioimage analysis, including the Crick BioImage Analysis Symposium 2021 and the “From Images to Knowledge (I2K) 2022” conference of the Center for Open Bioimage Analysis and Bioimaging North America.

The relative novelty of the Julia programming language, which is not yet an established industry standard, acted both as an obstacle and an incentive. On the downside, as fewer researchers are familiar with it than Python or R, it may seem harder to use than other bioinformatics software. However, this difficulty is in practice mitigated by a user-friendly syntax and by a good interoperability of the different languages. Dedicated libraries allow to execute foreign code in Julia (RCall.jl, PyCall.jl) or Julia code in other languages (JuliaCall, PyJulia). On the upside, more and more developers recognize the potential that Julia offers, and Julia is progressively becoming a sound and sustainable language choice for bioinformatics applications by combining performance and ease of use. There is a growing universe of Julia tools for the analysis of biological data and the budding developer community is very welcoming towards new packages (Roesch et al., 2021). Overall, it is relatively easy to interact with other package developers to ensure compatibility and integrate in the existing landscape. It is also easier to identify existing projects to avoid duplicates and to find an unoccupied niche for a new tool than in the crowded package ecosystem of more established languages.

This project did not end with the publication of the article and the package keeps maturing. Many more improvements are planned and under development. New features will include ways to make the methods more flexible and better deal with missing and non-numerical values. The performance could also be enhanced by supporting parallel computing for more steps, and avoiding storing complete datasets in memory via lazy loading, when possible. On the more conceptual side, the way we quantify distances between morphological profiles could and will be further improved, as detailed in subsection 4.5.

A last aspect that became evident while working on the analyses presented in the paper is the scarcity of benchmarking resources for imaging-based screens. This is made especially difficult by the variety of tasks performed in HCS experiments, but some open datasets are already freely available for common tasks such as mechanism of action prediction. Large collections of genetic and chemical perturbations without any “ground truth”, namely which morphological effects should cluster together, were also publicly shared. Two of the largest and most widely-used available collections are the Broad Bioimage Benchmark Collection (Ljosa et al., 2012) and the RxRx datasets from Recursion Pharmaceuticals, including recent cellular images of potential treatments for COVID-19 (Heiser et al., 2020). The Image Data Resource also aims at improving the situation by collecting reference imaging datasets along with extensive annotations allowing to connect the metadata to reference databases and ontologies (Williams et al., 2017).

4.2 Further profiling human lymphocytes

Our study of the immunological synapse (IS) of cytotoxic lymphocytes, presented in section 3.2, offered insight into how the actin cytoskeleton and the lytic granules are organized and how this structure is disrupted in the context of actinopathies. Simple machine learning models were able to precisely infer disease status and perturbations for cellular images based on morphological measurements. This supported the existence of a strong link between morphology and function, which could be more generally established in follow-up studies. We also observed that a static two-dimensional snapshot of the cells only provides an incomplete picture of the state of the IS, which we complemented by three-dimensional analyses of the synapse organization and live cell imaging. This is of prime importance when studying T and NK cells, as cytotoxicity is a dynamic process, and timelapse and dedicated assays are essential for obtaining a finer view of immune functions, for instance to study motility (Kamnev et al., 2021). Overall, the HCS experiments we conducted served as an efficient hypothesis-generating tool, giving a good overview of the spatial organization of our study object, which can then be completed by additional, more targeted assays.

In the future, it will also be important to work with models of immune activity that match more closely human physiology. Organoids have recently been used in image-based screens, showing that observing the interplay between cells adds considerable value to such experiments (Betge et al., 2019; Lukonin et al., 2020). Immune cells, which are individually motile, cannot be made to aggregate into three-dimensional structures, yet they are heavily influenced by interactions, both with immune cells and with other cell types and pathogens. The identification of infected cells as non-self drastically modulates the state of immune cells by triggering their activation. In low throughput cytotoxicity assays, the IS is typically observed by combining a population of cytotoxic lymphocytes with antigen-presenting or target cells, thus eliciting the formation of the synapse at their interface (Dupré et al., 2005; Kalinichenko et al., 2021; Mace & Orange, 2014; Salzer et al., 2016). Follow-up studies might explore whether it is also possible to adapt such co-culture methods to high-throughput assays. The interactions between host cells are shaping the immune response as well. The lymphocytes do not act in isolation, but are part of a complex system involving many cell types. This was well illustrated by Vladimer and colleagues, who screened a large chemical library on leukocytes from peripheral blood and observed that around 10% of the perturbations affected physical cell-cell interactions (Vladimer et al., 2017). Thus, they demonstrated that such interactions can be seen as a hallmark of immune response. Overall, our understanding of immunity is evolving rapidly, and more cell types have been recently shown to participate, including structural cells and red blood cells (Krausgruber et al., 2020; Lam et al., 2021). Integrating inter-cellular context, microenvironment and activation state of immune cells will therefore be a major challenge and opportunity to unravel the molecular basis of inborn errors of immunity, infections and cancers, and hopefully lead to novel treatments.

4.3 Understanding biological systems through perturbations

The efforts described in this thesis fall within our attempts to better characterize biological systems through their response to cellular perturbations. HCI offers a scalable and systematic way to quantify effects of genetic and chemical libraries. This allows us to look for principles underlying such responses. Accordingly, we strive to compile an arithmetic of cellular perturbations, permitting to predict the effect of combinatorial perturbations based on the knowledge of the biological system and of the effect of each individual perturbation in isolation. This approach was already adopted by our research group in a study of drug-drug interactions (Caldera et al., 2019). In this work, both structural and functional annotations of the compounds and the localization of their molecular targets on the interactome helped to contextualize the interactions. This project and the design of the perturbation arithmetic framework were discussed in depth in the doctoral thesis of Dr. Michael Caldera (Caldera, 2020).

Tackling the technological and analytical challenges inherent to HCI screens paves the way for larger and more systematic experiments. In addition to the published articles described above, I also worked on the design and analysis of large-scale screens, involving CRISPR knockouts in isolation or in combination with chemical treatments. This project was highly collaborative and involved researchers with different and complementary skill sets. It was conceptualized by Prof. Jörg Menche and me, with inputs from our collaborators. The group of Ass.-Prof. Joanna Loizou, and in particular Joana Ferreira da Silva, led the experimental work for the generation of the sgRNA library and shared their experience with the U-2-OS cell line. The group of Dr. Vanja Nagy, and in particular Christopher Fell and Viktoriia Kartysh, also participated in the generation of the library, brought their expertise with the SK-N-AS cell line and took care of the cell culture, transfection and seeding steps. The group of Dr. Loïc Dupré, and in particular Dr. Anton Kamnev, handled the staining and imaging of the cells and the preparation of the Jurkat cell line. The screening facility at CeMM, and in particular Anna Koren, assisted us with the preparation of the chemical library and the automated image acquisition on the PerkinElmer Operetta HCS platform. Data was also generated on the PerkinElmer Opera Phenix microscope from the group of Prof. Giulio Superti-Furga, which kindly gave us access. Finally, our group is in charge of all computational analyses. Raphael Bednarsky, supervised by Prof. Jörg Menche and myself, wrote his master’s thesis on an analysis of a subset of the genetic screens. I already analysed the data pertaining to chemical perturbations on their own, as presented in section 3.1. Finally, I am currently conducting all downstream data analyses across all remaining experiments.

These screening efforts are divided into two parts. First, we aimed to systematically explore the effect of intrinsic, genetic perturbations on cellular morphology, and the role of the biological context by using the same perturbations in different cell lines (Figure 4.1a). The first step was to design an sgRNA library which targets 240 genes. The selection of genetic targets was based, among other criteria, on their expected tendency to induce morphological changes, based on

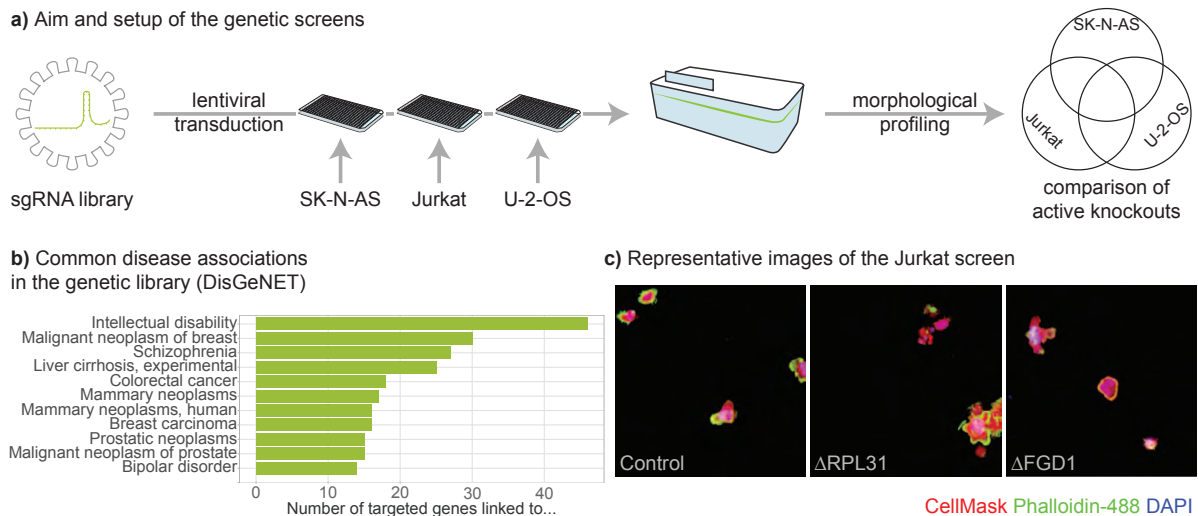


Figure 4.1: First part of the follow-up project dealing with genetic perturbations. **a)** Outline of the experiments. **b)** Most common disease associations among the 240 genes targeted in our experiment, according to the DisGeNET database (Pinero et al., 2015). **c)** Example images showing the morphological impact of RPL31 and FGD1 knockouts compared to a non-targeting control in Jurkat cells.

published data and simulations. The resulting library includes multiple Rho-GTPases and actin cytoskeleton remodelers, as they play an essential role in cell morphology (Svitkina, 2018) and genes related to several rare monogenic diseases, in particular inborn errors of immunity (Brigida et al., 2018; Dupré et al., 2015; Pfajfer et al., 2018) and neurodevelopmental disorders (Zamboni et al., 2018). Here as well, we benefit from the expertise of our experimental collaborators, who contribute their knowledge regarding gene editing, intellectual disabilities and inborn errors of immunity. The library, focusing on a limited number of genes, is practical and cost-efficient, and still covers multiple biological pathways and disease associations (Figure 4.1b).

Following its design, the library was synthesized and used to conduct genetic screens. Oligonucleotides were purchased to target four sites per gene. They were further inserted into pLCV3 vectors from Horizon Discovery, including the cas9 gene and a puromycin resistance gene. The four guides per gene were pooled, so that each well of three 96-well plates targets a given gene or includes a non-targeting control, and cloned into lentiviral particles. The neuroblastoma cell line SK-N-AS, the T lymphocyte line Jurkat and the osteosarcoma cell line U-2-OS were transduced in microplates with this library, and successfully transfected cells were selected with puromycin. The cells were then fixed, stained for DNA, beta-actin and total protein, and imaged on the PerkinElmer Opera Phenix automated microscope. The resulting images display distinct cellular morphology induced by some of the knockouts (Figure 4.1c).

The second part of the project is dedicated to the creation of the first map of chemical-genetic interactions in a human cell line. Thus, we are expanding the lab's previous work on combinatorial perturbations. Experimentally, this was achieved by systematically testing combinations of one chemical compound with one gene knockout from our custom sgRNA library. As for the assembly of the genetic library, compounds in the drug library were selected to maximize morphological activity and relevance in a broad biological context. We pursue

the hypothesis that basic principles of cellular response to perturbations are encoded in the interactome and that we can use cell morphology to observe and decipher these principles. Recent advances in the characterization of complex genetic interactions (Kuzmin et al., 2018) and drug-gene interactions (Piotrowski et al., 2017) were made in yeast solely based on growth assays. We believe that HCI has the potential to describe more precisely how different perturbations are affecting each other. Interactions between perturbations are typically defined as a deviation from the simple additive effect of the individual perturbations (Iorio et al., 2016). Our research group previously generalized this idea and expanded an existing computational framework (Caldera et al., 2019; Fischer et al., 2015) to allow for a detailed characterization of how intrinsic and extrinsic perturbations combine at the cell-level, for a well-defined and controlled model system. In this project, we tackle chemical-genetic interactions, representing extrinsic and intrinsic cellular perturbations, respectively, and aim to derive potential suppression (negative interaction), enhancement (positive interaction) and new effects that emerge as a result of the perturbation combination. We first knocked out genes in U-2-OS cells and plated them on drug plates (Figure 4.2a). We then acquired images using the PerkinElmer Operetta high-content analysis system, from which we extract quantitative descriptors of cellular morphology. In brief, we quantify the effect of two perturbations as vectors in a morphological space. An interaction then corresponds to a difference between the sum of the vectors of the individual perturbations and the measured vector of the combined perturbation (Figure 4.2b). Using this conceptual framework, we aim to identify and characterize the interactions in our model system. This goes beyond the traditional undirected interactions inferred from one-dimensional readout assays, such as cytotoxicity assays, and gives more insight about the molecular phenomena occurring, as morphology is an informative marker of cell function. We already collected the data and the investigations of interactions are ongoing. Some gene knockouts and compounds induce clear morphological changes (Figure 4.2c). In addition to investigating basic principles of gene-drug interactions, we envision that our work may contribute to advancing precision medicine. This could be done both by applying the principles uncovered, and through the validation of specific observations made in the screen. A careful selection of our experimental model system aims to expedite this ambition, while at the same time ensuring the more direct impact of this project on the biology of rare monogenic diseases. To this end, we chose to work with human cell lines and a selection of genes that are amenable to high-throughput imaging, as well as potentially relevant in the context of known genetic defects. Gene knockouts can indeed help to identify drug candidates in compound screens. For example, mutations in the *LMNA* gene that cause laminopathies can be rescued by remodelin, a small molecule targeting NAT10 (Larrieu et al., 2014).

Finally, by characterizing the obtained map of drug-gene interactions in a systematic manner, for example through network clusters and motifs, we hope to identify underlying rules contributing to the understanding of how intrinsic and extrinsic cell perturbations combine and affect the cell morphology. Overall, we expect to (i) gain new insights into the landscape of directed drug-gene interactions, (ii) expand our understanding of perturbations of molecular interactome

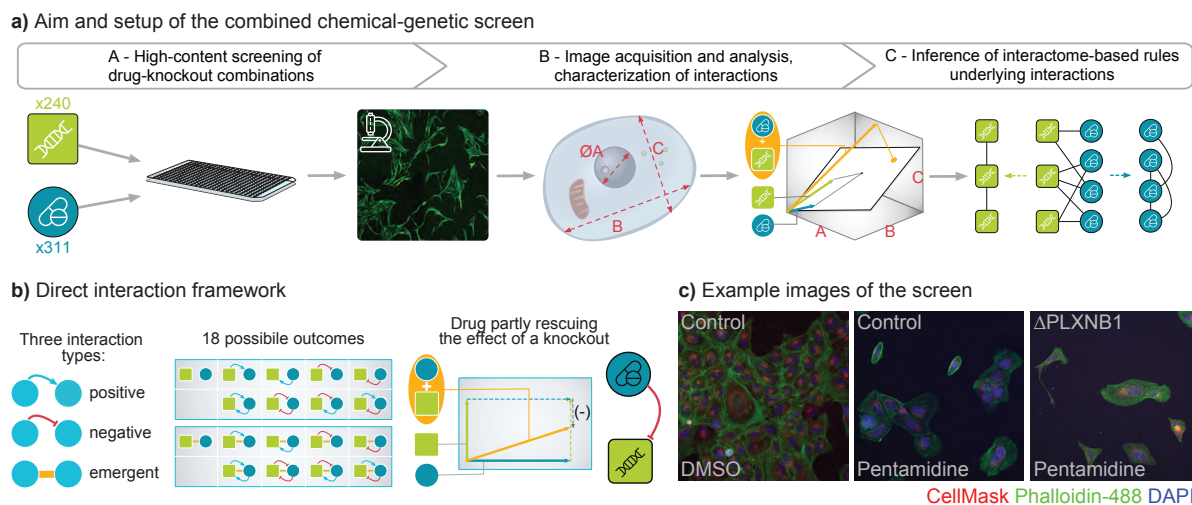


Figure 4.2: Second part of the follow-up project dealing with mixed chemical and genetic perturbations. **a)** Outline of the experiments. **b)** Interaction framework describing how two perturbations combine. Three types of interactions (left) can coexist and combine in 18 different outcomes (center). An example (right) shows how a reduced change in the morphological space in the direction spanned by the genetic perturbation could correspond to a drug rescue. **c)** Example images showing the effect of pentamidine with and without a knockout of PLXNB1, compared to DMSO-treated U-2-OS cells transduced with a non-targeting control sgRNA.

networks and (iii) identify core principles of how perturbations interact with each other that may also apply to other biological systems. This approach harnesses the power of systems biology and networks science to go beyond what can be concluded from the interactions individually, allowing for improved system-wide conclusions (Barabási et al., 2011), and embodies our vision that morphological profiling can work hand in hand with network medicine.

4.4 Expanding the reach of morphological profiling

Morphology is a complex phenotype linked to the global organization of the cell and resulting from the complex interplay between many actors, including organelles, tubulin and actin cytoskeletons. In this context, the phrase “Form follows function” is often used to emphasize that cell shape is a comprehensive readout that can be linked to cellular function, from migration to cell-cell interactions and structural properties of tissues (Caldera, 2020; Sailem et al., 2014). HCS can thus provide a more reliable marker of activity in many biological processes than molecular readouts, such as transcriptome- or proteome-level measurements, which are often poorly correlating with each other and weakly predict downstream effects including translation and signaling (Brunner et al., 2020; Gabor et al., 2021; Tuncbag et al., 2016). Thanks to a relatively low price per sample and an increasing availability of automated microscopes, HCS may serve as a reference multipurpose assay. It is able to provide single cell measurements, at a high-throughput and with high-dimensional readout, and includes spatial information that is

missing from many other assays. Yet, technologies based on sequencing and mass spectrometry tend to be favored for many applications. One of the reasons for this is certainly the difficulty to convert the raw images into usable and informative descriptions of the molecular changes that cells undergo. Conceivably, this may be eased more and more by advances in the field of morphological profiling. An overarching goal for HCS is therefore to reach the maturity level sufficient to be considered as an additional OMICS technology, to be used on its own or in integrative approaches.

The parallel development of other spatially-resolved assays could also blur the line between HCS and OMICS analyses. Crowned “method of the year” in 2020 by the journal *Nature Methods* (Marx, 2021), spatial transcriptomics combines gene expression and local information in tissue, at the resolution of a couple of cells. Common methods include FISH techniques, mentioned above in subsection 1.2.1, or involve massive parallel sequencing performed on spots of biological tissues after imaging, such as proposed by the popular Visium platform from 10x Genomics. Other emerging spatially-resolved technologies also include imaging mass cytometry, offering to simultaneously stain for tens of markers in pathology images, which already found applications ranging from breast cancer to COVID-19 (Jackson et al., 2020; Rendeiro et al., 2021), and multiplexed imaging, allowing for instance the combined detection of RNA and protein abundance (Schulz et al., 2018). Here as well, the data is best interpreted by integration across multiple modalities. For example, traditional histological images were used to increase the resolution of spatial transcriptomics experiments (Bergenstr hle et al., 2021; Pang et al., 2021). This suggests that methods to integrate spatially-resolved data with other data types will remain an important research direction.

Novel screening technologies also emerged, allowing the combined acquisition of morphological and genetic readouts, and often being high-throughput and suitable for large-scale genetic screens (Bock et al., 2022). Some methods allow for the physical isolation of cells with a morphological phenotype of choice (Schraivogel et al., 2022; Wheeler et al., 2020). Other methods rely on in-situ sequencing, where fluorescent barcodes are used to identify perturbations (Feldman et al., 2019; Reicher et al., 2020; Wang et al., 2019). If they become more popular, they could lead to significant changes in the subsequent analysis steps. For instance, comparisons between conditions would be made more direct, as there would no longer be plate layout effects. As for the arrayed setup proposed by de Groot and colleagues, which performs CRISPR editing in a subset of the imaged cells by low rate transient transfections, each image could include its own negative control in the form of unperturbed cells (de Groot et al., 2018). Real multimodal readouts could also be considered, allowing to simultaneously monitor the internal state of a cell at the transcriptional level, and the resulting morphology. This would provide a tool of choice to further refine the link between genotype and phenotype. However, the adoption of these methods to replace arrayed screens with more cost-effective alternatives will depend on the technological advances of pooled screens in general, as some issues such as their low sensitivity still remain.

As hinted by our ability to classify ARPC1B- and WASP-deficient patients, or by the recent

report of the prediction of lung cancer variants from HCS-derived features (Caicedo et al., 2021), it might be possible to differentiate between a limited number of genetic defects based on morphology alone. Yet, image-based profiles are likely to be best used not as a replacement of technologies describing the internal state of the cells but as a complementary view, closer to the function served by these cells. Independently from experimental setups providing multiple readouts, attempts to integrate and contextualize morphological information with other data sources are also becoming more common. For instance, studies put together morphological and transcriptomic profiles, derived from HCS and from the L1000 assay, to propose quantitative models of the relation between both feature types (Haghighi et al., 2021; Nassiri & McCall, 2018). The informativeness and reproducibility of both modalities were also extensively compared and their combination appeared beneficial (Way et al., 2021b).

4.5 Importance of distances to quantify morphological activity

In our analyses, we observed that the choice of a distance metric to quantify morphological activity has strong implications, and that the most suitable choice depends on specific requirements of the data at hand. HCI offers a rich readout which makes it possible to systematically describe differences compared to a baseline state, as with OMICS technologies. However, no general conclusion can be drawn when the effect or the sample size is not sufficient to guarantee that the observed effects are not entirely stochastic. Determining which perturbations induce statistically significant changes before characterizing such changes is therefore essential, yet sometimes neglected.

The mp-value approach, introduced above in subsection 1.2.6, proposes an empirical test to quantify whether the centers of two profile distributions, typically corresponding to a perturbation and a reference, differ significantly. We wanted a modern and computationally efficient implementation of such principles and specifically aimed to make the analyses unbiased by biological and technical outliers, as they are common in HCI data. To achieve these goals, we used robust statistics and non-linear dimensionality reduction, which is able to recapitulate effects that linear methods such as PCA would miss, in both studies compiled in this thesis. In our analyses, we quantified morphological distances in the low-dimensional space obtained after running UMAP (McInnes et al., 2018). In short, the method starts by building a fuzzy neighborhood graph by linking each data point to its closest neighbors, accounting for the uneven spatial distribution of the data. It then finds a projection of this graph to a low-dimensional Euclidean space that best preserves some of its characteristics. Ideally, one would keep the ability of the method to represent complex data structures in a computationally efficient way while avoiding the significant distortion induced by the graph layout step (Cooley et al., 2020). As mentioned in the paper describing BioProfiling.jl, working in the original space is often intractable and typically limited by the so-called curse of dimensionality. An

appealing alternative that our group is currently pursuing would be to measure distances in the neighborhood-based graph assembled by the UMAP algorithm while avoiding the subsequent bias-inducing projection step. Network distances could then be used to quantify changes between profiles for different conditions, now corresponding to distinct sets of nodes.

Whether on such a graph, in the original dataset, or in the reduced space, many different metrics can be used and are able to convey different properties of the data. In the analyses we conducted, I mainly quantified morphological distances with the Mahalanobis and the Hellinger distances. They differ in that the former compares a point to a distribution while the latter considers two distributions, although they both are measuring distances between two location estimators. This means that two profiles may have a null distance if they have the same center despite having different variances. It might be tempting to also compare the spread of these distributions and consider that a perturbation that induces an increase or a decrease in the heterogeneity of the profiles should be considered morphologically active. However, this would require to carefully consider what the profiles actually represent. Most assays differentiate between technical and biological replicates, when measurements are performed multiple times on the same sample, or on multiple samples, respectively (Blainey et al., 2014). Biological replicates are considered “true replicates”, as they increase the number of independent statistical sampling out of the biological distribution of interest. They are usually performed for secondary screens where the sensitivity of the test is important in order to confirm or invalidate previous findings, but not systematically conducted for primary screens (Malo et al., 2006). Technical replicates, on the other hand, are useful to reduce the variance in the sampling process and mitigate eventual experimental mistakes or artifacts, but fail to provide information about the variance of the biological process studied. HCS analyses constitute an analogous case. Screens often provide a large number of images per well, each displaying multiple cells, yet all measurements performed in the same well arise from a single treatment and cannot give insight into the robustness or heterogeneity of the perturbation. This has concrete implications regarding the interpretation of distances between profiles: When comparing profiles corresponding to technical replicates, such as multiple image-level profiles originating from the same well per condition, one first needs to ensure that the measurements are equally reproducible in all conditions. It is only then possible to attribute the observed variation in cell morphology to the effect of the perturbation. For instance, this could be an issue when a condition affects the cell count and the image-based profiles are aggregating cell-level measurements. Profiles would then show more variation in conditions affecting cell viability than in the conditions with a large number of cells. Additionally, when comparing profiles corresponding to biological replicates, a difference of spread could also correspond to varying responses to the perturbation, such as incomplete penetrance of a phenotype, varying genetic editing rates, or simply different outcomes of cell count bottlenecks. There are therefore multiple factors that may increase the spread of morphological profiles. Among those, the main factors that we presented are also illustrated on Figure 4.3. Disentangling all these putative causes of profile variation is a major challenge but has the potential to provide valuable insight about cellular response to perturbations.

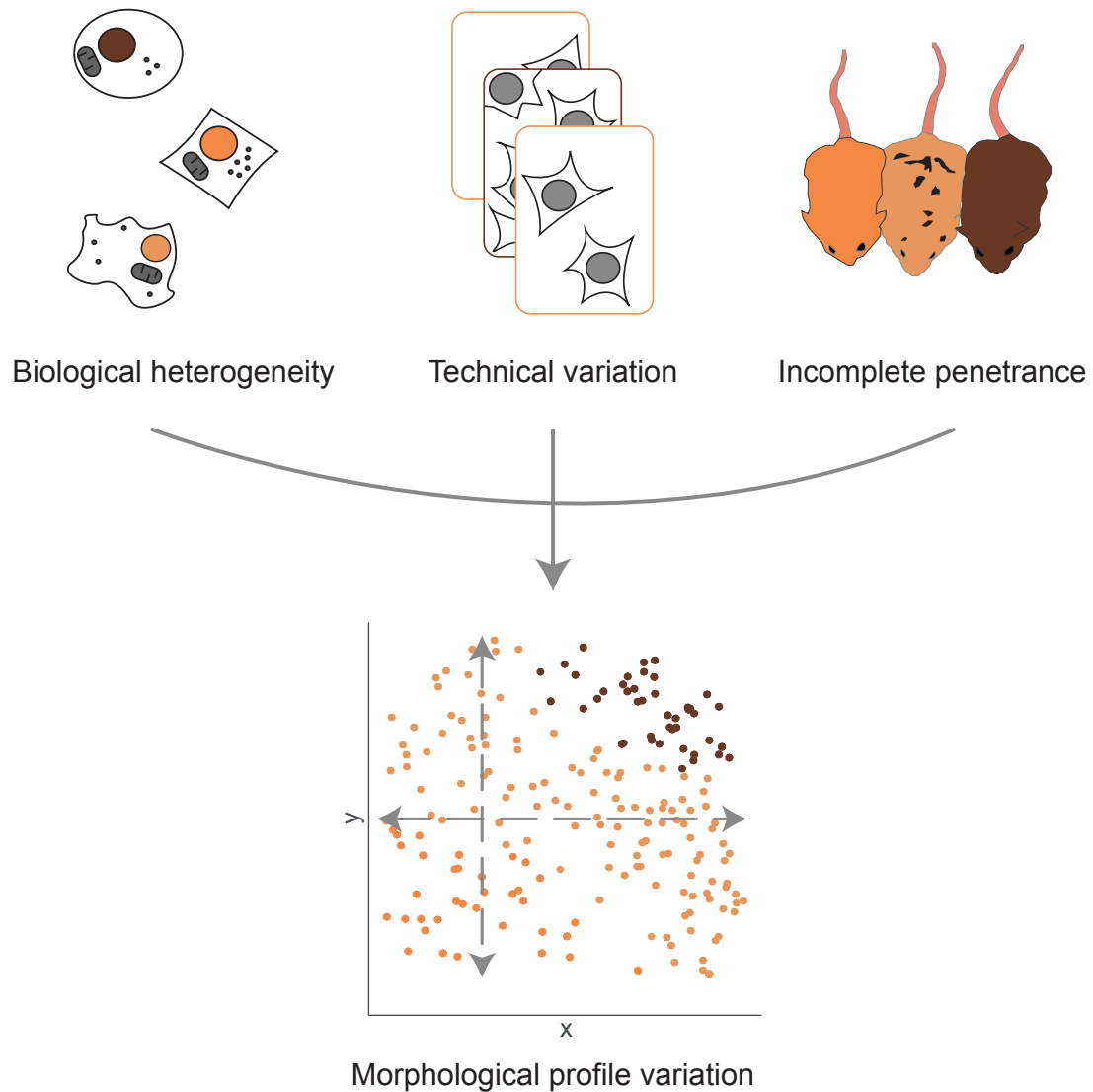


Figure 4.3: Schematic depiction of how different biological and technical parameters may affect the dispersion of morphological profiles. Biological heterogeneity corresponds to an actual variability in phenotype between samples. Technical variation covers all the changes that can be attributed to the experimental design and setup. Incomplete penetrance refers to the cases where a phenotypic change might only be visible in a fraction of the studied population. All these parameters may lead, in isolation or in combination, to the heterogeneity observed between measurements for a given condition. It is therefore necessary to implement proper experimental controls and replicates to unravel these effects and interpret profile heterogeneity.

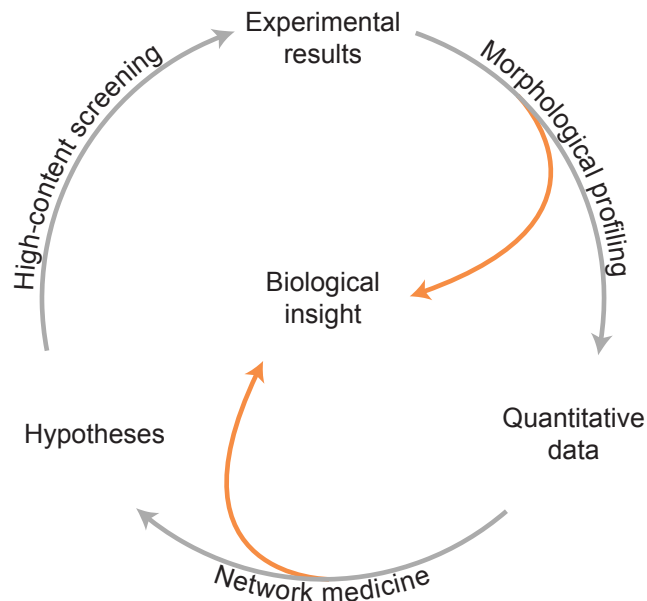


Figure 4.4: Interaction of high-content screening, morphological profiling and network medicine. We envision that the following “virtuous circle” may hold the key to the fast generation and testing of biological hypotheses. The proposed approach would combine emerging techniques from the field of high-content screening, morphological profiling and network medicine and benefit from a cross-disciplinary synergy bringing together experimental and computational biologists.

4.6 Interfacing morphological profiling and network medicine

Taken together, the manuscripts compiled in this thesis and the ongoing projects in our research group contribute to develop the link between morphological profiling and network medicine. Advances in the field of morphological profiling and new tools such as BioProfiling.jl make it increasingly more convenient and practical to assemble a comprehensive and quantitative description of changes occurring in image-based screens. However, this should only be seen as a means to an end, and should foster downstream analyses aiming to address major research questions and provide novel biological insight. To this end, the combination of morphological profiling and network medicine offers an appealing solution, as morphological profiling is dedicated to transforming raw imaging data into quantitative profiles that can be further interrogated, and network medicine is especially suitable to integrate and interpret well-curated data. On the one hand, network medicine proposes to generate hypotheses from existing data based on well-defined goals, such as how to target a protein, trigger a signaling cascade or explain the etiology of a disease. On the other hand, HCS and morphological profiling make it possible to quickly test such hypotheses experimentally, assess perturbation effects and guide follow-up assays. They embody how *in vitro* and *in silico* approaches may complement each other to advance our knowledge of biological systems. The potential synergy between HCI, morphological profiling and network medicine, as depicted in Figure 4.4, is supported by advances in these fields, including our own contributions. We conclude that such a virtuous circle holds the key to the

concrete implementation of the aims described in section 2 and in Figure 2.1, that is to test system-level hypotheses at a high-throughput and unlock the full information available from a screening readout.

A striking example is drug development, where independent approaches to prioritize and assess candidate drugs based on both morphology and molecular network properties were described, even for the same disease. Here, we may use COVID-19, the disease caused by the SARS-CoV-2 virus infection, as a study case for which many relevant articles were published in the span of two years, to address the challenges of the resulting pandemic. It was suggested at an early stage that existing data about other human coronaviruses integrated in PPI networks could support drug repurposing (Zhou et al., 2020). Rapidly, a large group of researchers formed a consortium to produce and share a high-confidence map of interactions between SARS-CoV-2 and human proteins, and found that some of them are druggable targets (Gordon et al., 2020). This valuable resource was later used by multiple studies, and notably to assess multiple network-based methods for drug repurposing (Morselli Gysi et al., 2021). Moreover, network approaches were used to integrate the various data available from multiple sources about all actors involved in the disease (Verstraete et al., 2020). The relevant scientific literature was also parsed and summarized into knowledge graphs specific of COVID-19 (Sharma et al., 2021). Independently, several groups developed screens for the unbiased image-based identification of approved drugs to revert morphological effects of SARS-CoV-2 infections in cell lines (Bakowski et al., 2021; Ellinger et al., 2021; Heiser et al., 2020), to study infection-induced platelet aggregation (Nishikawa et al., 2021) and to reduce the resulting fibrosis (Marwick et al., 2021). Of note, the latter study also used a network-based pathway analysis to explore the mechanism of action of its targets. The efforts made in describing the spatial organization of the disease were not limited to screens but also aimed to better understand the underlying mechanisms. For instance a landscape of lung pathology was assembled using imaging mass cytometry (Rendeiro et al., 2021). HCS methods, albeit limited to the analysis of single intensity parameters, were also able to distinguish between healthy and infected cell populations (Francis et al., 2021) or to support the maintained activity of drugs for multiple variants of concern (Vangeel et al., 2022).

All in all, the amount of articles describing HCS and network medicine methods to tackle the problems caused by the emerging infectious disease, and the prominence of this body of work, are remarkable. This demonstrates nicely the flexibility and the many promises of these approaches, as well as their success within the research community. Nevertheless, it should be noted that no novel treatment was approved as a direct result of these publications so far, which emphasizes how challenging the clinical translation of biomedical research is. However, these strategies are maturing rapidly, among others within the framework of precision medicine. Methods delivering an informed decision based on molecular measurements of a patient are now reaching outside of academic research, and are making good progress regarding drug development for multiple diseases. For instance, a platform proposing the ex-vivo prioritization of treatments in patients with hematological cancers based on an image-based functional assay underwent a successful prospective study (NCT03096821) and is now followed-up by a larger-scope ran-

domized study (NCT04470947) (Kornauth et al., 2021). The technology used in this trial is also developed commercially by Alcyte, a company recently acquired by Exscientia, which itself focuses on modern computational approaches to drug discovery and is supporting a phase 1 trial (NCT04727138). Recursion Pharmaceuticals, a company specialized on the automated analysis of cellular morphology, is conducting a phase 2 clinical trial for cerebral cavernous malformation (NCT05085561), following previous work done by its founders (Gibson et al., 2015). The field of network medicine also explored new treatment options, notably with applications to oncology (Nogales et al., 2021). System-level insight guided the development of preclinical studies for drug repurposing aiming to ease the testing of *in silico* approaches (Schmidt, 2018). Of note, a study investigated repurposed drugs to treat glioblastoma in a phase 1 and 2 trial (NCT02770378), and justified the efficacy of its protocol based on the altered signaling and metabolic networks observed in patients (Kast et al., 2014). Overall, we hope to see more contributions of modern computational approaches integrated with high-throughput and high-content experiments to both basic biology and medical applications in the future.

Bibliography

- Ahmed, S., Zhou, Z., Zhou, J., & Chen, S.-Q. (2016). Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine. *Genomics, Proteomics & Bioinformatics*, 14(5), 298–313.
URL <https://linkinghub.elsevier.com/retrieve/pii/S1672022916301413>
- Algar, W. R., Hildebrandt, N., Vogel, S. S., & Medintz, I. L. (2019). FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. *Nature Methods*, 16(9), 815–829.
URL <http://www.nature.com/articles/s41592-019-0530-8>
- Apweiler, R., Beissbarth, T., Berthold, M. R., Blüthgen, N., Burmeister, Y., Dammann, O., Deutsch, A., Feuerhake, F., Franke, A., Hasenauer, J., Hoffmann, S., Höfer, T., Jansen, P. L., Kaderali, L., Klingmüller, U., Koch, I., Kohlbacher, O., Kuepfer, L., Lammert, F., Maier, D., Pfeifer, N., Radde, N., Rehm, M., Roeder, I., Saez-Rodriguez, J., Sax, U., Schmeck, B., Schuppert, A., Seilheimer, B., Theis, F. J., Vera, J., & Wolkenhauer, O. (2018). Whither systems medicine? *Experimental & Molecular Medicine*, 50(3), e453.
URL <http://www.nature.com/doifinder/10.1038/emm.2017.290>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29.
URL <http://www.ncbi.nlm.nih.gov/pubmed/10802651><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3037419><http://www.nature.com/doifinder/10.1038/75556>
- Bai, S. W., Herrera-Abreu, M. T., Rohn, J. L., Racine, V., Tajadura, V., Suryavanshi, N., Bechtel, S., Wiemann, S., Baum, B., & Ridley, A. J. (2011). Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biology*, 9(1), 54.
URL <https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-9-54>

- Bakowski, M. A., Beutler, N., Wolff, K. C., Kirkpatrick, M. G., Chen, E., Nguyen, T.-T. H., Riva, L., Shaabani, N., Parren, M., Ricketts, J., Gupta, A. K., Pan, K., Kuo, P., Fuller, M., Garcia, E., Teijaro, J. R., Yang, L., Sahoo, D., Chi, V., Huang, E., Vargas, N., Roberts, A. J., Das, S., Ghosh, P., Woods, A. K., Joseph, S. B., Hull, M. V., Schultz, P. G., Burton, D. R., Chatterjee, A. K., McNamara, C. W., & Rogers, T. F. (2021). Drug repurposing screens identify chemical entities for the development of COVID-19 interventions. *Nature Communications*, 12(1), 3309.
URL <http://www.nature.com/articles/s41467-021-23328-0>
- Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 16878.
URL <http://www.nature.com/articles/s41598-017-17204-5>
- Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature reviews. Genetics*, 12(1), 56–68.
URL <http://www.ncbi.nlm.nih.gov/pubmed/21164525><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3140052>
- Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101–113.
URL <http://www.nature.com/doifinder/10.1038/nrg1272>
- Barker, C. G., Petsalaki, E., Giudice, G., Sero, J., Ekpenyong, E. N., Bakal, C., & Petsalaki, E. (2022). Identification of phenotype-specific networks from paired gene expression-cell shape imaging data. *Genome Research*, (p. gr.276059.121).
URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.276059.121>
- Barnes, E. C., Kumar, R., & Davis, R. A. (2016). The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. *Natural Product Reports*, 33(3), 372–381.
URL <http://xlink.rsc.org/?DOI=C5NP00121H>
- Bengio, Y., Courville, A., & Vincent, P. (2012). Representation Learning: A Review and New Perspectives. *arXiv*.
URL <http://arxiv.org/abs/1206.5538>
- Bergenstråhle, L., He, B., Bergenstråhle, J., Abalo, X., Mirzazadeh, R., Thrane, K., Ji, A. L., Andersson, A., Larsson, L., Stakenborg, N., Boeckxstaens, G., Khavari, P., Zou, J., Lundeberg, J., & Maaskola, J. (2021). Super-resolved spatial transcriptomics by deep data fusion. *Nature Biotechnology*.
URL <https://www.nature.com/articles/s41587-021-01075-3>
- Bergthaler, A., & Menche, J. (2017). The immune system as a social network. *Nature Immunology*, 18(5), 481–482.
URL <http://www.nature.com/articles/ni.3727>
- Bernas, T. (2005). Basics of Digital Microscopy. *Current Protocols in Cytometry*, 31(1).
URL <https://onlinelibrary.wiley.com/doi/10.1002/0471142956.cy1202s31>
- Betge, J., Rindtorff, N., Sauer, J., Rauscher, B., Dingert, C., Gaitantzi, H., Herweck, F., Miersch, T., Valentini, E., Hauber, V., Gutting, T., Frank, L., Belle, S., Gaiser, T., Buchholz, I.,

- Jesenofsky, R., Härtel, N., Zhan, T., Fischer, B., Breitkopf-Heinlein, K., Burgermeister, E., Ebert, M. P., & Boutros, M. (2019). Multiparametric phenotyping of compound effects on patient derived organoids. *bioRxiv*, (p. 660993).
URL <https://www.biorxiv.org/node/763693.full>
- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1), 65–98.
URL <https://epubs.siam.org/doi/10.1137/141000671>
- Billmann, M., Horn, T., Fischer, B., Sandmann, T., Huber, W., & Boutros, M. (2016). A genetic interaction map of cell cycle regulators. *Molecular biology of the cell*, 27(8), 1397–407.
URL <http://www.ncbi.nlm.nih.gov/pubmed/26912791><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4831891>
- Blainey, P., Krzywinski, M., & Altman, N. (2014). Replication. *Nature Methods*, 11(9), 879–880.
URL <http://www.nature.com/articles/nmeth.3091>
- Blais, E. M., Rawls, K. D., Dougherty, B. V., Li, Z. I., Kolling, G. L., Ye, P., Wallqvist, A., & Papin, J. A. (2017). Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nature Communications*, 8, 14250.
URL <http://www.nature.com/doiifinder/10.1038/ncomms14250>
- Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B., Stanley, G., Chen, S., Garnett, M., Li, W., Moffat, J., Qi, L. S., Shapiro, R. S., Shendure, J., Weissman, J. S., & Zhuang, X. (2022). High-content CRISPR screening. *Nature Reviews Methods Primers*, 2(1), 8.
URL <https://www.nature.com/articles/s43586-021-00093-4>
- Bossi, A., & Lehner, B. (2009). Tissue specificity and the human protein interaction network. *Molecular Systems Biology*, 5(1), 260.
URL <https://onlinelibrary.wiley.com/doi/10.1038/msb.2009.17>
- Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, 33(S3), 228–237.
URL <http://www.nature.com/articles/ng1090z>
- Bougen-Zhukov, N., Loh, S. Y., Lee, H. K., & Loo, L.-H. (2017). Large-scale image-based screening and profiling of cellular phenotypes. *Cytometry Part A*, 91(2), 115–125.
URL <http://doi.wiley.com/10.1002/cyto.a.22909>
- Boutros, M., Brás, L. P., & Huber, W. (2006). Analysis of cell-based RNAi screens. *Genome Biology*, 7(7), R66.
URL <https://doi.org/10.1186/gb-2006-7-7-r66>
- Boutros, M., Heigwer, F., & Laufer, C. (2015). Microscopy-Based High-Content Screening. *Cell*, 163(6), 1314–1325.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867415014877>
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7), 1177–1186.

- URL <http://www.ncbi.nlm.nih.gov/pubmed/28622505><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5536862>
- Bray, M.-A., Gustafsdottir, S. M., Rohban, M. H., Singh, S., Ljosa, V., Sokolnicki, K. L., Bittker, J. A., Bodycombe, N. E., Dančák, V., Hasaka, T. P., Hon, C. S., Kemp, M. M., Li, K., Walpita, D., Wawer, M. J., Golub, T. R., Schreiber, S. L., Clemons, P. A., Shamji, A. F., & Carpenter, A. E. (2017). A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *GigaScience*, 6(12), 1–5.
URL <http://academic.oup.com/gigascience/article/6/12/1/2865213>
- Bray, M.-A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S. M., Gibson, C. C., & Carpenter, A. E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature Protocols*, 11(9), 1757–1774.
URL <http://www.nature.com/doifinder/10.1038/nprot.2016.105>
- Breinig, M., Klein, F. A., Huber, W., & Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Molecular Systems Biology*, 11(12), 846–846.
URL <http://msb.embopress.org/cgi/doi/10.15252/msb.20156400>
- Brigida, I., Zoccolillo, M., Cicalese, M. P., Pfajfer, L., Barzaghi, F., Scala, S., Oleaga-Quintas, C., Álvarez-Álvarez, J. A., Sereni, L., Giannelli, S., Sartirana, C., Dionisio, F., Pavesi, L., Benavides-Nieto, M., Basso-Ricci, L., Capasso, P., Mazzi, B., Rosain, J., Marcus, N., Lee, Y. N., Somech, R., Degano, M., Raiola, G., Caorsi, R., Picco, P., Moncada Velez, M., Khourieh, J., Arias, A. A., Bousfiha, A., Issekutz, T., Issekutz, A., Boisson, B., Dobbs, K., Villa, A., Lombardo, A., Neven, B., Moshous, D., Casanova, J.-L., Franco, J. L., Notarangelo, L. D., Scielzo, C., Volpi, S., Dupré, L., Bustamante, J., Gattorno, M., & Aiuti, A. (2018). T-cell defects in patients with ARPC1B germline mutations account for combined immunodeficiency. *Blood*, 132(22), 2362–2374.
URL <https://ashpublications.org/blood/article/132/22/2362/107711/Tcell-defects-in-patients-with-ARPC1B-germline>
- Bruch, P.-M., Giles, H. A. R., Kolb, C., Herbst, S. A., Becirovic, T., Roider, T., Lu, J., Scheinost, S., Wagner, L., Huellein, J., Berest, I., Kriegsmann, M., Kriegsmann, K., Zgorzelski, C., Dreger, P., Zaugg, J. B., Müller-Tidow, C., Zenz, T., Huber, W., & Dietrich, S. (2021). Combinatorial drug-microenvironment interaction mapping reveals cell-extrinsic drug resistance mechanisms and clinically relevant patient subgroups in CLL. *bioRxiv*, (p. 2021.07.23.453514).
URL <http://biorxiv.org/content/early/2021/11/18/2021.07.23.453514.abstract>
- Brummelkamp, T. R., Bernards, R., & Agami, R. (2002). A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science*, 296(5567), 550–553.
URL <https://www.science.org/doi/10.1126/science.1068999>
- Brunner, J. S., Vulliard, L., Hofmann, M., Kieler, M., Lercher, A., Vogel, A., Russier, M., Brüggenthies, J. B., Kerndl, M., Saferding, V., Niederreiter, B., Junza, A., Frauenstein, A., Scholtyssek, C., Mikami, Y., Klavins, K., Krönke, G., Bergthaler, A., O’Shea, J. J., Weichhart, T., Meissner, F., Smolen, J. S., Cheng, P., Yanes, O., Menche, J., Murray, P. J., Sharif, O., Blüml, S., & Schabbauer, G. (2020). Environmental arginine controls multinuclear giant cell

- metabolism and formation. *Nature Communications*, 11(1), 431.
URL <http://www.nature.com/articles/s41467-020-14285-1>
- Bryce, N. S., Failes, T. W., Stehn, J. R., Baker, K., Zahler, S., Arzhaeva, Y., Bischof, L., Lyons, C., Dedova, I., Arndt, G. M., Gaus, K., Goult, B. T., Hardeman, E. C., Gunning, P. W., & Lock, J. G. (2019). High-Content Imaging of Unbiased Chemical Perturbations Reveals that the Phenotypic Plasticity of the Actin Cytoskeleton Is Constrained. *Cell Systems*, 9(5), 496–507.
- Buphamalai, P., Kokotovic, T., Nagy, V., & Menche, J. (2021). Network analysis reveals rare disease signatures across multiple levels of biological organization. *Nature Communications*, 12(1), 6306.
URL <https://www.nature.com/articles/s41467-021-26674-1>
- Caicedo, J. C., Arevalo, J., Piccioni, F., Bray, M.-A., Hartland, C. L., Wu, X., Brooks, A. N., Berger, A. H., Boehm, J. S., Carpenter, A. E., & Singh, S. (2021). Cell Painting predicts impact of lung cancer variants. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/11/20/2021.11.18.469171>
- Caicedo, J. C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A. S., Barry, J. D., Bansal, H. S., Kraus, O., Wawer, M., Paavolainen, L., Herrmann, M. D., Rohban, M., Hung, J., Hennig, H., Concannon, J., Smith, I., Clemons, P. A., Singh, S., Rees, P., Horvath, P., Lington, R. G., & Carpenter, A. E. (2017). Data-analysis strategies for image-based cell profiling. *Nature Methods*, 14(9), 849–863.
URL <http://www.ncbi.nlm.nih.gov/pubmed/28858338><http://www.nature.com/doifinder/10.1038/nmeth.4397>
- Caicedo, J. C., McQuin, C., Goodman, A., Singh, S., & Carpenter, A. E. (2018). Weakly Supervised Learning of Single-Cell Feature Embeddings. *bioRxiv*, (p. 293431).
URL <https://www.biorxiv.org/content/10.1101/293431v1>
- Caldera, M. (2020). *Mapping a directed interaction network of drug perturbations using a high-dimensional cell morphology readout*. Ph.D. thesis, Medizinische Universität Wien.
- Caldera, M., Müller, F., Kaltenbrunner, I., Licciardello, M. P., Lardeau, C.-H., Kubicek, S., & Menche, J. (2019). Mapping the perturbome network of cellular perturbations. *Nature Communications*, 10(1), 5140.
URL <http://www.nature.com/articles/s41467-019-13058-9>
- Campbell, K. S., & Hasegawa, J. (2013). Natural killer cell biology: An update and future directions. *Journal of Allergy and Clinical Immunology*, 132(3), 536–544.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0091674913010658>
- Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I., Friman, O., Guertin, D. A., Chang, J., Lindquist, R. A., Moffat, J., Golland, P., & Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10), R100.
URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-10-r100>
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green Fluorescent Protein as a Marker for Gene Expression. *Science*, 263(5148), 802–805.
URL <https://www.science.org/doi/10.1126/science.8303295>

- Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D., & Carpenter, A. E. (2020). Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nature Reviews Drug Discovery*.
URL <http://www.nature.com/articles/s41573-020-00117-w>
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*(6233).
URL <https://www.science.org/doi/10.1126/science.aaa6090>
- Cheng, F., Desai, R. J., Handy, D. E., Wang, R., Schneeweiss, S., Barabási, A.-L., & Loscalzo, J. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature communications*, *9*(1), 2691.
URL <http://www.ncbi.nlm.nih.gov/pubmed/30002366><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6043492>
- Cheng, F., Kovács, I. A., & Barabási, A.-L. (2019). Network-based prediction of drug combinations. *Nature Communications*, *10*(1), 1197.
URL <http://www.nature.com/articles/s41467-019-09186-x>
- Cheng, Y., Glaeser, R. M., & Nogales, E. (2017). How Cryo-EM Became so Hot. *Cell*, *171*(6), 1229–1231.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867417313259>
- Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y. L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., Huss, M., Soh, B.-S., Kraus, P., Li, P., Lufkin, T., Lim, B., Clarke, N. D., Bard, F., & Ng, H.-H. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, *468*(7321), 316–320.
URL <http://www.nature.com/articles/nature09531>
- Cooley, S. M., Hamilton, T., Ray, J. C. J., & Deeds, E. J. (2020). A novel metric reveals previously unrecognized distortion in dimensionality reduction of scRNA-Seq data. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2020/09/28/689851>
- Coster, A. D., Wichaidit, C., Rajaram, S., Altschuler, S. J., & Wu, L. F. (2014). A simple image correction method for high-throughput microscopy. *Nature Methods*, *11*(6), 602–602.
URL <http://www.nature.com/articles/nmeth.2971>
- Datlinger, P., Rendeiro, A. F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L. C., Kuchler, A., Alpar, D., & Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods*, *14*(3), 297–301.
URL <http://www.nature.com/articles/nmeth.4177>
- de Boer, P., Hoogenboom, J. P., & Giepmans, B. N. G. (2015). Correlated light and electron microscopy: ultrastructure lights up! *Nature Methods*, *12*(6), 503–513.
URL <http://www.nature.com/articles/nmeth.3400>
- de Groot, R., Lüthi, J., Lindsay, H., Holtackers, R., & Pelkmans, L. (2018). Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Molecular Systems Biology*, *14*(1), e8064.
URL <http://msb.embopress.org/lookup/doi/10.15252/msb.20178064>
- de Wet, T. J., Winkler, K. R., Mhlanga, M., Mizrahi, V., & Warner, D. F. (2020). Arrayed CRISPRi and quantitative imaging describe the morphotypic landscape of essential

- mycobacterial genes. *eLife*, 9.
URL <https://elifesciences.org/articles/60083>
- Dinstag, G., & Shamir, R. (2020). PRODIGY: personalized prioritization of driver genes. *Bioinformatics (Oxford, England)*.
- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S., Friedman, N., & Regev, A. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, 167(7), 1853–1866.
URL <http://www.ncbi.nlm.nih.gov/pubmed/27984732><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5181115>
- Doench, J. G. (2017). Am I ready for CRISPR? A user’s guide to genetic screens. *Nature Reviews Genetics*, 19(2), 67–80.
URL <http://www.nature.com/doifinder/10.1038/nrg.2017.97>
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184–191.
URL <http://www.nature.com/articles/nbt.3437>
- Dranoff, G. (2004). Cytokines in cancer pathogenesis and cancer therapy. *Nature Reviews Cancer*, 4(1), 11–22.
URL <http://www.nature.com/articles/nrc1252>
- Dupré, L., Aiuti, A., Trifari, S., Martino, S., Saracco, P., Bordignon, C., & Roncarolo, M.-G. (2002). Wiskott-Aldrich Syndrome Protein Regulates Lipid Raft Dynamics during Immunological Synapse Formation. *Immunity*, 17(2), 157–166.
URL <https://linkinghub.elsevier.com/retrieve/pii/S1074761302003606>
- Dupré, L., Andolfi, G., Tangye, S. G., Clementi, R., Locatelli, F., Aricò, M., Aiuti, A., & Roncarolo, M.-G. (2005). SAP controls the cytolytic activity of CD8+ T cells against EBV-infected cells. *Blood*, 105(11), 4383–4389.
URL <https://ashpublications.org/blood/article/105/11/4383/19502/SAP-controls-the-cytolytic-activity-of-CD8-T-cells>
- Dupré, L., Houmadi, R., Tang, C., & Rey-Barroso, J. (2015). T Lymphocyte Migration: An Action Movie Starring the Actin and Associated Actors. *Frontiers in Immunology*, 6, 586.
URL <http://journal.frontiersin.org/Article/10.3389/fimmu.2015.00586/abstract>
- Ellinger, B., Bojkova, D., Zaliani, A., Cinatl, J., Claussen, C., Westhaus, S., Keminer, O., Reinshagen, J., Kuzikov, M., Wolf, M., Geisslinger, G., Gribbon, P., & Ciesek, S. (2021). A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. *Scientific Data*, 8(1), 70.
URL <http://www.nature.com/articles/s41597-021-00848-4>
- Elmarakeby, H. A., Hwang, J., Arafah, R., Crowdis, J., Gang, S., Liu, D., AlDubayan, S. H., Salari, K., Kregel, S., Richter, C., Arnoff, T. E., Park, J., Hahn, W. C., & Van Allen, E. M. (2021). Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880), 348–352.
URL <https://www.nature.com/articles/s41586-021-03922-4>

- Erni, R., Rossell, M. D., Kisielowski, C., & Dahmen, U. (2009). Atomic-Resolution Imaging with a Sub-50-pm Electron Probe. *Physical Review Letters*, 102(9), 096101.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.102.096101>
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., & D'Eustachio, P. (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Research*, 44(D1), D481–D487.
URL <http://www.ncbi.nlm.nih.gov/pubmed/26656494><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4702931><https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1351>
- Falk, T., Mai, D., Bensch, R., Çiçek, , Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T., & Ronneberger, O. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1), 67–70.
URL <http://www.nature.com/articles/s41592-018-0261-2>
- Feldman, D., Singh, A., Schmid-Burgk, J. L., Carlson, R. J., Mezger, A., Garrity, A. J., Zhang, F., & Blainey, P. C. (2019). Optical Pooled Screens in Human Cells. *Cell*, 179(3), 787–799.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867419310670>
- Feldman, I., Rzhetsky, A., & Vitkup, D. (2008). Network properties of genes harboring inherited disease mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), 4323–8.
URL <http://www.ncbi.nlm.nih.gov/pubmed/18326631><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2393821>
- Fell, C. W., Hagelkruys, A., Cicvaric, A., Horrer, M., Liu, L., Li, J. S. S., Stadlmann, J., Polyansky, A. A., Mereiter, S., Tejada, M. A., Kokotović, T., Scaramuzza, A., Twyman, K. A., Morrow, M. M., Juusola, J., Yan, H., Wang, J., Burmeister, M., Andersen, T. L., Wirsberger, G., Holmskov, U., Perrimon, N., Zagrović, B., Monje, F. J., Moeller, J. B., Penninger, J. M., & Nagy, V. (2021). FIBCD1 is a Conserved Receptor for Chondroitin Sulphate Proteoglycans of the Brain Extracellular Matrix and a Candidate Gene for a Complex Neurodevelopmental Disorder. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/09/10/2021.09.09.459581>
- Ferreira da Silva, J., Meyenberg, M., & Loizou, J. I. (2021). Tissue specificity of DNA repair: the CRISPR compass. *Trends in Genetics*, 37(11), 958–962.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0168952521002006>
- Fischer, B., Sandmann, T., Horn, T., Billmann, M., Chaudhary, V., Huber, W., & Boutros, M. (2015). A map of directional genetic interactions in a metazoan cell. *eLife*, 4.
URL <https://elifesciences.org/articles/05464>
- Fortelny, N., & Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biology*, 21(1), 190.
URL <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02100-5>

- Francis, R., Le Bideau, M., Jardot, P., Grimaldier, C., Raoult, D., Bou Khalil, J. Y., & La Scola, B. (2021). High-speed large-scale automated isolation of SARS-CoV-2 from clinical samples using miniaturized co-culture coupled to high-content screening. *Clinical Microbiology and Infection*, 27(1), 1–128.
URL <https://linkinghub.elsevier.com/retrieve/pii/S1198743X2030570X>
- Frederick, M., & Tukey, J. (1977). *Data analysis and regression : a second course in statistics*. Reading, Mass: Addison-Wesley Pub. Co.
- Gabor, A., Tognetti, M., Driessen, A., Tanevski, J., Guo, B., Cao, W., Shen, H., Yu, T., Chung, V., Bodenmiller, B., Saez-Rodriguez, J., Prusokas, A., Prusokas, A., Retkute, R., Rajasekar, A., Raman, K., Sudhakar, M., Rengaswamy, R., Shih, E. S., Kim, M., Cho, C., Kim, D., Oh, H., Hwang, J., Jongtae, K., Nam, Y., Yoon, S., Kwon, T., Lee, K., Chaudhary, S., Sharma, N., Bande, S., Cankut Cubuk, G. G. f. z., Gundogdu, P., Dopazo, J., Rian, K., Loucera, C., Falco, M. M., Garrido-Rodriguez, M., Peña-Chilet, M., Chen, H., Turu, G., Hunyadi, L., Misak, A., Guo, B., Cao, W., Shen, H., Zhou, L., Jiang, X., Zhang, P., Rai, A., Kutum, R., Rana, S., Srinivasan, R., Pradhan, S., Li, J., Bajic, V., Van Neste, C., Barradas-bautista, D., Albarade, S. A., Nikolskiy, I., Sinkala, M., Tran, D., Nguyen, H., Nguyen, T., Wu, A., DeMeo, B., Hie, B., Singh, R., Liu, J., Chen, X., Saiz, L., Vilar, J. M. G., Qiu, P., Gosain, A., Dhall, A., Bajaj, D., Kaur, H., Bagaria, K., Chauhan, M., Sharma, N., Raghava, G., Patiyal, S., Hao, J., Peng, J., Ning, S., Ma, Y., Wei, Z., Aalto, A., Goncalves, J., Mombaerts, L., Dai, X., Zheng, J., Mundra, P., Xu, F., Wang, J., Kant Singh, K., & Lee, M. (2021). Cell-to-cell and type-to-type heterogeneity of signaling networks: insights from the crowd. *Molecular Systems Biology*, 17(10).
URL <https://onlinelibrary.wiley.com/doi/10.15252/msb.202110402>
- Gao, J., Li, D., & Havlin, S. (2014). From a single network to a network of networks. *National Science Review*, 1(3), 346–356.
URL <https://academic.oup.com/nsr/article/1/3/346/2460684>
- Gapp, B. V., Konopka, T., Penz, T., Dalal, V., Bürckstümmer, T., Bock, C., & Nijman, S. M. (2016). Parallel reverse genetic screening in mutant human cells using transcriptomics. *Molecular Systems Biology*, 12(8), 879.
URL <http://msb.embopress.org/lookup/doi/10.15252/msb.20166890>
- German, Y. (2020). *High content cell imaging reveals actin cytoskeleton-mediated control of the immunological synapse*. Ph.D. thesis, Université Toulouse 3 - Paul Sabatier.
- Ghiassian, S. D., Menche, J., & Barabási, A.-L. (2015). A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. *PLOS Computational Biology*, 11(4), e1004120.
URL <http://dx.plos.org/10.1371/journal.pcbi.1004120>
- Gibson, C. C., Zhu, W., Davis, C. T., Bowman-Kirigin, J. A., Chan, A. C., Ling, J., Walker, A. E., Goitre, L., Delle Monache, S., Retta, S. F., Shiu, Y.-T. E., Grossmann, A. H., Thomas, K. R., Donato, A. J., Lesniewski, L. A., Whitehead, K. J., & Li, D. Y. (2015). Strategy for Identifying Repurposed Drugs for the Treatment of Cerebral Cavernous Malformation. *Circulation*, 131(3), 289–299.
URL <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.114.010403>
- Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., &

- Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 159(3), 647–661.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867414011787>
- Giuliano, K. A., DeBiasio, R. L., Dunlay, R. T., Gough, A., Volosky, J. M., Zock, J., Pavlakis, G. N., & Taylor, D. L. (1997). High-Content Screening: A New Approach to Easing Key Bottlenecks in the Drug Discovery Process. *Journal of Biomolecular Screening*, 2(4), 249–259.
URL <http://journals.sagepub.com/doi/10.1177/108705719700200410>
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21), 8685–90.
URL <http://www.ncbi.nlm.nih.gov/pubmed/17502601>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1885563>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351.
URL <http://www.nature.com/articles/nrg.2016.49>
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., Haas, P., Polacco, B. J., Braberg, H., Fabius, J. M., Eckhardt, M., Soucheray, M., Bennett, M. J., Cakir, M., McGregor, M. J., Li, Q., Meyer, B., Roesch, F., Vallet, T., Mac Kain, A., Miorin, L., Moreno, E., Naing, Z. Z. C., Zhou, Y., Peng, S., Shi, Y., Zhang, Z., Shen, W., Kirby, I. T., Melnyk, J. E., Chiorba, J. S., Lou, K., Dai, S. A., Barrio-Hernandez, I., Memon, D., Hernandez-Armenta, C., Lyu, J., Mathy, C. J. P., Perica, T., Pilla, K. B., Ganesan, S. J., Saltzberg, D. J., Rakesh, R., Liu, X., Rosenthal, S. B., Calviello, L., Venkataramanan, S., Liboy-Lugo, J., Lin, Y., Huang, X.-P., Liu, Y., Wankowicz, S. A., Bohn, M., Safari, M., Ugur, F. S., Koh, C., Savar, N. S., Tran, Q. D., Shengjuler, D., Fletcher, S. J., O'Neal, M. C., Cai, Y., Chang, J. C. J., Broadhurst, D. J., Klippsten, S., Sharp, P. P., Wenzell, N. A., Kuzuoglu-Ozturk, D., Wang, H.-Y., Trenker, R., Young, J. M., Cavero, D. A., Hiatt, J., Roth, T. L., Rathore, U., Subramanian, A., Noack, J., Hubert, M., Stroud, R. M., Frankel, A. D., Rosenberg, O. S., Verba, K. A., Agard, D. A., Ott, M., Emerman, M., Jura, N., von Zastrow, M., Verdin, E., Ashworth, A., Schwartz, O., D'Enfert, C., Mukherjee, S., Jacobson, M., Malik, H. S., Fujimori, D. G., Ideker, T., Craik, C. S., Floor, S. N., Fraser, J. S., Gross, J. D., Sali, A., Roth, B. L., Ruggero, D., Taunton, J., Kortemme, T., Beltrao, P., Vignuzzi, M., García-Sastre, A., Shokat, K. M., Shoichet, B. K., & Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583(7816), 459–468.
URL <http://www.nature.com/articles/s41586-020-2286-9>
- Greenwald, N. F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C. C., McIntosh, B. J., Leow, K. X., Schwartz, M. S., Pavelchek, C., Cui, S., Camplisson, I., Bar-Tal, O., Singh, J., Fong, M., Chaudhry, G., Abraham, Z., Moseley, J., Warshawsky, S., Soon, E., Greenbaum, S., Risom, T., Hollmann, T., Bendall, S. C., Keren, L., Graf, W., Angelo, M., & Van Valen, D. (2021). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*.
URL <https://www.nature.com/articles/s41587-021-01094-0>
- Guney, E., Menche, J., Vidal, M., & Barabási, A.-L. (2016). Network-based in silico drug efficacy

- screening. *Nature Communications*, 7(1), 10331.
URL <http://www.nature.com/articles/ncomms10331>
- Gustafsdottir, S., Ljosa, V., Sokolnicki, K., Walpita, D., Kemp, M., Petri Seiler, K., Carrel, H., Golub, T., Schreiber, S., Clemons, P., Carpenter, A., & Shamji, A. (2013). Multiplex Cytological Profiling Assay to Measure Diverse Cellular States. *PLoS ONE*.
- Haghighi, M., Singh, S., Caicedo, J., & Carpenter, A. (2021). High-Dimensional Gene Expression and Morphology Profiles of Cells across 28,000 Genetic and Chemical Perturbations. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/09/08/2021.09.08.459417>
- Haglund, C. M., & Welch, M. D. (2011). Pathogens and polymers: Microbe–host interactions illuminate the cytoskeleton. *The Journal of Cell Biology*, 195(1), 7–17.
URL <http://www.ncbi.nlm.nih.gov/pubmed/21969466><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3187711><http://www.jcb.org/lookup/doi/10.1083/jcb.201103148>
- Han, K., Pierce, S. E., Li, A., Spees, K., Anderson, G. R., Seoane, J. A., Lo, Y.-H., Dubreuil, M., Olivas, M., Kamber, R. A., Wainberg, M., Kostyrko, K., Kelly, M. R., Yousefi, M., Simpkins, S. W., Yao, D., Lee, K., Kuo, C. J., Jackson, P. K., Sweet-Cordero, A., Kundaje, A., Gentles, A. J., Curtis, C., Winslow, M. M., & Bassik, M. C. (2020). CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature*, (pp. 1–6).
URL <http://www.nature.com/articles/s41586-020-2099-x>
- Hancock, J., & Menche, J. (2021). Structure and Function in Complex Biological Networks. In *Systems Medicine*, (pp. 8–16). Elsevier.
URL <https://linkinghub.elsevier.com/retrieve/pii/B9780128012383116393>
- Heigwer, F., Scheeder, C., Miersch, T., Schmitt, B., Blass, C., Pour Jamnani, M. V., & Boutros, M. (2018). Time-resolved mapping of genetic interactions to model rewiring of signaling pathways. *eLife*, 7.
URL <http://www.ncbi.nlm.nih.gov/pubmed/30592458><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6319608><https://elifesciences.org/articles/40174>
- Heiser, K., McLean, P. F., Davis, C. T., Fogelson, B., Gordon, H. B., Jacobson, P., Hurst, B., Miller, B., Alfa, R. W., Earnshaw, B. A., Victors, M. L., Chong, Y. T., Haque, I. S., Low, A. S., & Gibson, C. C. (2020). Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2020/04/23/2020.04.21.054387>
- Hogan, A., Blomqvist, E., Cochez, M., D’amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022). Knowledge Graphs. *ACM Computing Surveys*, 54(4), 1–37.
URL <https://dl.acm.org/doi/10.1145/3447772>
- Hu, Y., Chen, C.-h., Ding, Y.-y., Wen, X., Wang, B., Gao, L., & Tan, K. (2019). Optimal control nodes in disease-perturbed networks as targets for combination therapy. *Nature Communications*, 10(1), 2180.
URL <http://www.nature.com/articles/s41467-019-10215-y>

- Hubert, M., & Debruyne, M. (2009). Breakdown value. *WIREs Computational Statistics*, 1(3), 296–302.
URL <https://onlinelibrary.wiley.com/doi/10.1002/wics.34>
- Hutz, J. E., Nelson, T., Wu, H., McAllister, G., Moutsatsos, I., Jaeger, S. A., Bandyopadhyay, S., Nigsch, F., Cornett, B., Jenkins, J. L., & Selinger, D. W. (2013). The Multidimensional Perturbation Value. *Journal of Biomolecular Screening*, 18(4), 367–377.
URL <http://journals.sagepub.com/doi/10.1177/1087057112469257>
- Ideker, T., Galitski, T., & Hood, L. (2001). A NEW APPROACH TO DECODING LIFE: Systems Biology. *Annual Review of Genomics and Human Genetics*, 2(1), 343–372.
URL <https://www.annualreviews.org/doi/10.1146/annurev.genom.2.1.343>
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., van Dyk, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R. K., Liu, Q., Mironenko, T., Mitropoulos, X., Richardson, L., Wang, J., Zhang, T., Moran, S., Sayols, S., Soleimani, M., Tamborero, D., Lopez-Bigas, N., Ross-Macdonald, P., Esteller, M., Gray, N. S., Haber, D. A., Stratton, M. R., Benes, C. H., Wessels, L. F. A., Saez-Rodriguez, J., McDermott, U., & Garnett, M. J. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, 166(3), 740–754.
URL <http://www.ncbi.nlm.nih.gov/pubmed/27397505><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4967469>
- Jackson, H. W., Fischer, J. R., Zanotelli, V. R., Ali, H. R., Mechera, R., Soysal, S. D., Moch, H., Muenst, S., Varga, Z., Weber, W. P., & Bodenmiller, B. (2020). The single-cell pathology landscape of breast cancer. *Nature*, 578(7796), 615–620.
- Ji, N., Shroff, H., Zhong, H., & Betzig, E. (2008). Advances in the speed and resolution of light microscopy. *Current Opinion in Neurobiology*, 18(6), 605–616.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0959438809000166>
- Jiao, Y., Ahmed, U., Sim, M. M., Bejar, A., Zhang, X., Talukder, M. M. U., Rice, R., Flannick, J., Podgornaia, A. I., Reilly, D. F., Engreitz, J. M., Kost-Alimova, M., Hartland, K., Mercader, J.-M., Georges, S., Wagh, V., Tadin-Strapps, M., Doench, J. G., Edwardson, J. M., Rochford, J. J., Rosen, E. D., & Majithia, A. R. (2019). Discovering metabolic disease gene interactions by correlated effects on cellular morphology. *Molecular Metabolism*, 24, 108–119.
URL <https://linkinghub.elsevier.com/retrieve/pii/S2212877819301188>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816–821.
URL <https://www.science.org/doi/10.1126/science.1225829>
- Jonas, E., & Kording, K. P. (2017). Could a Neuroscientist Understand a Microprocessor? *PLOS Computational Biology*, 13(1), e1005268.
URL <https://dx.plos.org/10.1371/journal.pcbi.1005268>
- Jones, T. R., Carpenter, A., & Golland, P. (2005). Voronoi-Based Segmentation of Cells on Image Manifolds. In *Proceedings of the First International Conference on Computer Vision for Biomedical Image Applications*, CVBIA'05, (p. 535–543). Berlin, Heidelberg: Springer-Verlag.
URL https://doi.org/10.1007/11569541_54

- Kalinichenko, A., Perinetti Casoni, G., Dupré, L., Trotta, L., Huemer, J., Galgano, D., German, Y., Haladik, B., Pazmandi, J., Thian, M., Yüce Petronczki, , Chiang, S. C., Taskinen, M., Hekkala, A., Kauppila, S., Lindgren, O., Tapiainen, T., Kraakman, M. J., Vettenranta, K., Lomakin, A. J., Saarela, J., Seppänen, M. R. J., Bryceson, Y. T., & Boztug, K. (2021). RhoG deficiency abrogates cytotoxicity of human lymphocytes and causes hemophagocytic lymphohistiocytosis. *Blood*, 137(15), 2033–2045.
URL <https://ashpublications.org/blood/article/137/15/2033/475089/RhoG-deficiency-abrogates-cytotoxicity-of-human>
- Kamnev, A., Lacouture, C., Fusaro, M., & Dupré, L. (2021). Molecular Tuning of Actin Dynamics in Leukocyte Migration as Revealed by Immune-Related Actinopathies. *Frontiers in Immunology*, 12.
URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.750537/full>
- Kampmann, M. (2018). CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chemical Biology*, 13(2), 406–416.
URL <https://pubs.acs.org/doi/10.1021/acscchembio.7b00657>
- Kang, J., Hsu, C.-H., Wu, Q., Liu, S., Coster, A. D., Posner, B. A., Altschuler, S. J., & Wu, L. F. (2015). Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. *Nature Biotechnology*, 34(1), 70–77.
URL <http://www.nature.com/doifinder/10.1038/nbt.3419>
- Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease.
URL <https://www-nature-com.ez.srv.meduniwien.ac.at/articles/nrg.2018.4.pdf>
- Kast, R. E., Karpel-Massler, G., & Halatsch, M.-E. (2014). CUSP9* treatment protocol for recurrent glioblastoma: aprepitant, artesunate, auranofin, captopril, celecoxib, disulfiram, itraconazole, ritonavir, sertraline augmenting continuous low dose temozolomide. *Oncotarget*, 5(18), 8052–8082.
URL <https://www.oncotarget.com/lookup/doi/10.18632/oncotarget.2408>
- Kawakubo, H., Matsui, Y., Kushima, I., Ozaki, N., & Shimamura, T. (2019). A network of networks approach for modeling interconnected brain tissue-specific networks. *Bioinformatics*, 35(17), 3092–3101.
URL <https://academic.oup.com/bioinformatics/article/35/17/3092/5289537>
- Kensert, A., Harrison, P. J., & Spjuth, O. (2019). Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes. *SLAS DISCOVERY: Advancing Life Sciences R&D*, (p. 247255521881875).
URL <http://journals.sagepub.com/doi/10.1177/2472555218818756>
- Khawatmi, M., Steux, Y., Zourob, S., & Sailem, H. (2021). ShapoGraphy: a glyph-oriented visualisation approach for creating pictorial representations of bioimaging data. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/04/09/2021.04.07.438792>
- Kim, H. S., Lee, K., Kim, S. J., Cho, S., Shin, H. J., Kim, C., & Kim, J. S. (2018). Arrayed CRISPR screen with image-based assay reliably uncovers host genes required for coxsackievirus infection. *Genome Research*.
- Kim, M., Park, J., Bouhaddou, M., Kim, K., Rojc, A., Modak, M., Soucheray, M., McGregor, M. J., O’Leary, P., Wolf, D., Stevenson, E., Foo, T. K., Mitchell, D., Herrington, K. A., Muñoz, D. P., Tutuncuoglu, B., Chen, K.-H., Zheng, F., Kreisberg, J. F., Diolaiti, M. E.,

- Gordan, J. D., Coppé, J.-P., Swaney, D. L., Xia, B., van 't Veer, L., Ashworth, A., Ideker, T., & Krogan, N. J. (2021). A protein interaction landscape of breast cancer. *Science*, 374 (6563). URL <https://www.science.org/doi/10.1126/science.abf3066>
- Kitsak, M., Sharma, A., Menche, J., Guney, E., Ghiassian, S. D., Loscalzo, J., & Barabási, A.-L. (2016). Tissue Specificity of Human Disease Module. *Scientific Reports*. URL <http://www.nature.com/articles/srep35241.pdf>
- Kornauth, C., Pemovska, T., Vladimer, G. I., Bayer, G., Bergmann, M., Eder, S., Eichner, R., Erl, M., Esterbauer, H., Exner, R., Felsleitner-Hauer, V., Forte, M., Gaiger, A., Geissler, K., Greinix, H. T., Gstöttner, W., Hacker, M., Hartmann, B. L., Hauswirth, A. W., Heinemann, T., Heintel, D., Hoda, M. A., Hopfinger, G., Jaeger, U., Kazianka, L., Kenner, L., Kiesewetter, B., Krall, N., Krajnik, G., Kubicek, S., Le, T., Lubowitzki, S., Mayerhoefer, M. E., Menschel, E., Merkel, O., Miura, K., Müllauer, L., Neumeister, P., Noesslinger, T., Ocko, K., Öhler, L., Panny, M., Pichler, A., Porpaczy, E., Prager, G. W., Raderer, M., Ristl, R., Ruckser, R., Salamon, J., Schiefer, A.-I., Schmolke, A.-S., Schwarzing, I., Selzer, E., Sillaber, C., Skrabs, C., Sperr, W. R., Srndic, I., Thalhammer, R., Valent, P., van der Kouwe, E., Vanura, K., Vogt, S., Waldstein, C., Wolf, D., Zielinski, C. C., Zojer, N., Simonitsch-Klupp, I., Superti-Furga, G., Snijder, B., & Staber, P. B. (2021). Functional Precision Medicine Provides Clinical Benefit in Advanced Aggressive Hematologic Cancers and Identifies Exceptional Responders. *Cancer Discovery*. URL <http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-21-0538>
- Kraus, O. Z., Grys, B. T., Ba, J., Chong, Y., Frey, B. J., Boone, C., & Andrews, B. J. (2017). Automated analysis of high-content microscopy data with deep learning. *Molecular systems biology*, 13(4), 924. URL <http://www.ncbi.nlm.nih.gov/pubmed/28420678><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5408780>
- Krausgruber, T., Fortelny, N., Fife-Gernedl, V., Senekowitsch, M., Schuster, L. C., Lercher, A., Nemc, A., Schmidl, C., Rendeiro, A. F., Bergthaler, A., & Bock, C. (2020). Structural cells are key regulators of organ-specific immune responses. *Nature*, 583(7815), 296–302. URL <http://www.nature.com/articles/s41586-020-2424-4>
- Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Usaj, M. M., Leeuwen, J. v., Koch, E. N., Pons, C., Dagilis, A. J., Pryszlak, M., Wang, J. Z. Y., Hanchard, J., Riggi, M., Xu, K., Heydari, H., Luis, B.-J. S., Shuteriqi, E., Zhu, H., Dyk, N. V., Sharifpoor, S., Costanzo, M., Loewith, R., Caudy, A., Bolnick, D., Brown, G. W., Andrews, B. J., Boone, C., & Myers, C. L. (2018). Systematic analysis of complex genetic interactions. *Science*, 360(6386), eaao1729. URL <http://science.sciencemag.org.ez.srv.meduniwien.ac.at/content/360/6386/eaao1729>
- Lam, L. K. M., Murphy, S., Kokkinaki, D., Venosa, A., Sherrill-Mix, S., Casu, C., Rivella, S., Weiner, A., Park, J., Shin, S., Vaughan, A. E., Hahn, B. H., Odom John, A. R., Meyer, N. J., Hunter, C. A., Worthen, G. S., & Mangalmurti, N. S. (2021). DNA binding to TLR9 expressed by red blood cells promotes innate immune activation and anemia. *Science Translational Medicine*, 13(616). URL <https://www.science.org/doi/10.1126/scitranslmed.abj1008>

- Laureau, D., Britton, S., Demir, M., Rodriguez, R., & Jackson, S. P. (2014). Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science (New York, N.Y.)*, 344(6183), 527–32.
URL <http://www.ncbi.nlm.nih.gov/pubmed/24786082><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4246063>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15(2), R29.
URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29>
- Lazebnik, Y. (2002). Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer cell*, 2(3), 179–82.
URL <http://www.ncbi.nlm.nih.gov/pubmed/12242150>
- Lee, A. Y., St.Onge, R. P., Proctor, M. J., Wallace, I. M., Nile, A. H., Spagnuolo, P. A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M. K., Cheung-Ong, K., Torres, N. P., Spear, E. D., Han, M. K. L., Schlecht, U., Suresh, S., Duby, G., Heisler, L. E., Surendra, A., Fung, E., Urbanus, M. L., Gebbia, M., Lissina, E., Miranda, M., Chiang, J. H., Aparicio, A. M., Zeghouf, M., Davis, R. W., Cherfils, J., Boutry, M., Kaiser, C. A., Cummins, C. L., Trimble, W. S., Brown, G. W., Schimmer, A. D., Bankaitis, V. A., Nislow, C., Bader, G. D., & Giaever, G. (2014). Mapping the Cellular Response to Small Molecules Using Chemogenomic Fitness Signatures. *Science*, 344(6180), 208–211.
URL <http://science.sciencemag.org.ez.srv.meduniwien.ac.at/content/sci/344/6180/208.full.pdf><https://www.science.org/doi/10.1126/science.1250217>
- Lichtman, J. W., & Conchello, J.-A. (2005). Fluorescence microscopy. *Nature Methods*, 2(12), 910–919.
URL <http://www.nature.com/articles/nmeth817>
- Lin, S., Liu, H., Svenningsen, E. B., Wollesen, M., Jacobsen, K. M., Andersen, F. D., Moyano-Villameriel, J., Pedersen, C. N., Nørby, P., Tørring, T., & Poulsen, T. B. (2021). Expanding the antibacterial selectivity of polyether ionophore antibiotics through diversity-focused semisynthesis. *Nature Chemistry*, 13(1), 47–55.
URL <http://www.nature.com/articles/s41557-020-00601-1>
- Liu, C., Zhao, J., Lu, W., Dai, Y., Hockings, J., Zhou, Y., Nussinov, R., Eng, C., & Cheng, F. (2020). Individualized genetic network analysis reveals new therapeutic vulnerabilities in 6,700 cancer genomes. *PLoS Computational Biology*, 16(2), e1007701.
- Liu, W., Wang, J., Wang, T., & Xie, H. (2014). Construction and Analyses of Human Large-Scale Tissue Specific Networks. *PLoS ONE*, 9(12), e115074.
URL <https://dx.plos.org/10.1371/journal.pone.0115074>
- Ljosa, V., Sokolnicki, K. L., & Carpenter, A. E. (2012). Annotated high-throughput microscopy image sets for validation. *Nature Methods*, 9(7), 637.
URL <http://www.nature.com/articles/nmeth.2083>
- López-Otín, C., & Kroemer, G. (2021). Hallmarks of Health. *Cell*, 184(1), 33–63.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867420316068>
- Loscalzo, J., Barabási, A.-L., & Silverman, E. K. (2017). *Network medicine: complex systems in human disease and therapeutics*. Harvard University Press.

- Lu, A. X., Kraus, O. Z., Cooper, S., & Moses, A. M. (2019). Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLOS Computational Biology*, 15(9), e1007348.
URL <http://dx.plos.org/10.1371/journal.pcbi.1007348>
- Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charlotiaux, B., Choi, D., Côté, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., Knapp, J. J., Kovács, I. A., Lemmens, I., Mee, M. W., Mellor, J. C., Pollis, C., Pons, C., Richardson, A. D., Schlabach, S., Teeking, B., Yadav, A., Babor, M., Balcha, D., Basha, O., Bowman-Colin, C., Chin, S.-F., Choi, S. G., Colabella, C., Coppin, G., D'Amata, C., De Ridder, D., De Rouck, S., Duran-Frigola, M., Ennajdaoui, H., Goebels, F., Goehring, L., Gopal, A., Haddad, G., Hatchi, E., Helmy, M., Jacob, Y., Kassa, Y., Landini, S., Li, R., van Lieshout, N., MacWilliams, A., Markey, D., Paulson, J. N., Rangarajan, S., Rasla, J., Rayhan, A., Rolland, T., San-Miguel, A., Shen, Y., Sheykhkarimli, D., Sheynkman, G. M., Simonovsky, E., Taşan, M., Tejeda, A., Tropepe, V., Twizere, J.-C., Wang, Y., Weatheritt, R. J., Weile, J., Xia, Y., Yang, X., Yeger-Lotem, E., Zhong, Q., Aloy, P., Bader, G. D., De Las Rivas, J., Gaudet, S., Hao, T., Rak, J., Tavernier, J., Hill, D. E., Vidal, M., Roth, F. P., & Calderwood, M. A. (2020). A reference map of the human binary protein interactome. *Nature*, 580(7803), 402–408.
URL <http://www.nature.com/articles/s41586-020-2188-x>
- Lukonin, I., Serra, D., Challet Meylan, L., Volkmann, K., Baaten, J., Zhao, R., Meeusen, S., Colman, K., Maurer, F., Stadler, M. B., Jenkins, J., & Liberali, P. (2020). Phenotypic landscape of intestinal organoid regeneration. *Nature*, 586(7828), 275–280.
URL <http://www.nature.com/articles/s41586-020-2776-9>
- Lüscher Dias, T., Schuch, V., Beltrão-Braga, P. C. B., Martins-de Souza, D., Brentani, H. P., Franco, G. R., & Nakaya, H. I. (2020). Drug repositioning for psychiatric and neurological disorders through a network medicine approach. *Translational Psychiatry*, 10(1), 141.
URL <http://www.nature.com/articles/s41398-020-0827-5>
- Lyubimova, A., Itzkovitz, S., Junker, J. P., Fan, Z. P., Wu, X., & van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. *Nature Protocols*, 8(9), 1743–1758.
URL <http://www.nature.com/articles/nprot.2013.109>
- Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., & Ideker, T. (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods*, 15(4), 290–298.
URL <http://www.nature.com/articles/nmeth.4627>
- Mace, E. M., & Orange, J. S. (2014). Lytic immune synapse function requires filamentous actin deconstruction by Coronin 1A. *Proceedings of the National Academy of Sciences*, 111(18), 6708–6713.
URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1314975111>
- Magger, O., Waldman, Y. Y., Ruppín, E., & Sharan, R. (2012). Enhancing the Prioritization of Disease-Causing Genes through Tissue Specific Protein Interaction Networks. *PLoS Computational Biology*, 8(9), e1002690.
URL <https://dx.plos.org/10.1371/journal.pcbi.1002690>
- Mahdessian, D., Cesnik, A. J., Gnann, C., Danielsson, F., Stenström, L., Arif, M., Zhang, C.,

- Le, T., Johansson, F., Shutten, R., Bäckström, A., Axelsson, U., Thul, P., Cho, N. H., Carja, O., Uhlén, M., Mardinoglu, A., Stadler, C., Lindskog, C., Ayoglu, B., Leonetti, M. D., Pontén, F., Sullivan, D. P., & Lundberg, E. (2021). Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature*, 590(7847), 649–654.
URL <http://www.nature.com/articles/s41586-021-03232-9>
- Malo, N., Hanley, J. A., Cerquozzi, S., Pelletier, J., & Nadon, R. (2006). Statistical practice in high-throughput screening data analysis. *Nature Biotechnology*, 24(2), 167–175.
URL <http://www.nature.com/articles/nbt1186>
- Malpica, N., de Solórzano, C. O., Vaquero, J. J., Santos, A., Vallcorba, I., García-Sagredo, J. M., & del Pozo, F. (1998). Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, 28(4), 289–297.
URL [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0320\(19970801\)28:4%3C289::AID-CYT03%3E3.0.CO;2-7](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0320(19970801)28:4%3C289::AID-CYT03%3E3.0.CO;2-7)
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753.
URL <http://www.nature.com/articles/nature08494>
- Marini Thian, Birgit Hoeger, Anton Kamnev, Fiona Poyer, Sevgi Köstel Bal, Michael Caldera, Raúl Jiménez-Heredia, Jakob Huemer, Winfried F. Pickl, Miriam Groß, Stephan Ehl, Carrie L. Lucas, Jörg Menche, Caroline Hutter, Andishe Attarbaschi, Loïc Dupré, & Kaan Boztug (2020). Germline biallelic PIK3CG mutations in a multifaceted immunodeficiency with immune dysregulation. *Haematologica*, 105(10), e488.
URL <https://haematologica.org/article/view/9635>
- Maron, B. A., Wang, R.-S., Shevtsov, S., Drakos, S. G., Arons, E., Wever-Pinzon, O., Huggins, G. S., Samokhin, A. O., Oldham, W. M., Aguib, Y., Yacoub, M. H., Rowin, E. J., Maron, B. J., Maron, M. S., & Loscalzo, J. (2021). Individualized interactomes for network-based precision medicine in hypertrophic cardiomyopathy with implications for other clinical pathophenotypes. *Nature Communications*, 12(1), 873.
URL <http://www.nature.com/articles/s41467-021-21146-y>
- Marwick, J. A., Elliott, R. J. R., Longden, J., Makda, A., Hirani, N., Dhaliwal, K., Dawson, J. C., & Carragher, N. O. (2021). Application of a High-Content Screening Assay Utilizing Primary Human Lung Fibroblasts to Identify Antifibrotic Drugs for Rapid Repurposing in COVID-19 Patients. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, 26(9), 1091–1106.
URL <http://journals.sagepub.com/doi/10.1177/24725552211019405>
- Marx, V. (2021). Method of the Year: spatially resolved transcriptomics. *Nature Methods*, 18(1), 9–14.
URL <http://www.nature.com/articles/s41592-020-01033-y>
- Mattiazzi Usaj, M., Sahin, N., Friesen, H., Pons, C., Usaj, M., Masinas, M. P. D., Shuteriqi, E., Shkurin, A., Aloy, P., Morris, Q., Boone, C., & Andrews, B. J. (2020). Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability.

- Molecular Systems Biology*, 16(2), 1–27.
URL <https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20199243>
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861.
URL <https://doi.org/10.21105/joss.00861>
- McQuin, C., Goodman, A., Chernyshev, V., Kametsky, L., Cimini, B. A., Karhohs, K. W., Doan, M., Ding, L., Rafelski, S. M., Thirstrup, D., Wiegraebe, W., Singh, S., Becker, T., Caicedo, J. C., & Carpenter, A. E. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biology*, 16(7), e2005970.
URL <http://dx.plos.org/10.1371/journal.pbio.2005970>
- Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A., & Olivo-Marin, J.-C. (2016). Imagining the future of bioimage analysis. *Nature Biotechnology*, 34(12), 1250–1255.
URL <http://www.nature.com/doifinder/10.1038/nbt.3722>
- Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., Johnson, R., Segrè, A. V., Djebali, S., Niarchou, A., Wright, F. A., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E. T., Ardlie, K. G., & Guigó, R. (2015). The human transcriptome across tissues and individuals. *Science*.
- Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A. L. (2015). Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224), 841.
URL <http://www.ncbi.nlm.nih.gov/pubmed/25700523><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4435741><http://www.sciencemag.org/cgi/doi/10.1126/science.1257601>
- Meyer, M. J., Beltrán, J. F., Liang, S., Fragoza, R., Rumack, A., Liang, J., Wei, X., & Yu, H. (2018). Interactome INSIDER: a structural interactome browser for genomic studies. *Nature Methods*, 15(2), 107–114.
URL <http://www.nature.com/doifinder/10.1038/nmeth.4540>
- Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepper, A. M., Hinkle, G., Piqani, B., Eisenhaure, T. M., Luo, B., Grenier, J. K., Carpenter, A. E., Foo, S. Y., Stewart, S. A., Stockwell, B. R., Hacohen, N., Hahn, W. C., Lander, E. S., Sabatini, D. M., & Root, D. E. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell*, 124(6), 1283–1298.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867406002388>
- Mondal, S., Hegarty, E., Martin, C., Gökçe, S. K., Ghorashian, N., & Ben-Yakar, A. (2016). Large-scale microfluidics providing high-resolution and high-throughput screening of *Caenorhabditis elegans* poly-glutamine aggregation model. *Nature Communications*, 7(1), 13023.
URL <http://www.nature.com/articles/ncomms13023>
- Morselli Gysi, D., do Valle, , Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S. D., Patten, J. J., Davey, R. A., Loscalzo, J., & Barabási, A.-L. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proceedings of the National Academy of Sciences*, 118(19), e2025581118.
URL <http://www.pnas.org/lookup/doi/10.1073/pnas.2025581118>

- Mullard, A. (2019). Machine learning brings cell imaging promises into focus. *Nature Reviews Drug Discovery*, 18(9), 653–655.
URL <http://www.nature.com/articles/d41573-019-00144-2>
- Muthuswamy, S. K., & Xue, B. (2012). Cell Polarity as a Regulator of Cancer Cell Behavior Plasticity. *Annual Review of Cell and Developmental Biology*, 28(1), 599–625.
URL <http://www.ncbi.nlm.nih.gov/pubmed/22881459><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3997262><http://www.annualreviews.org/doi/10.1146/annurev-cellbio-092910-154244>
- Nassiri, I., & McCall, M. N. (2018). Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. *Nucleic Acids Research*, 46(19), e116–e116.
URL <https://academic.oup.com/nar/article/46/19/e116/5054089>
- Neumann, B., Walter, T., Hériché, J.-K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U., Cetin, C., Sieckmann, F., Pau, G., Kabbe, R., Wünsche, A., Satagopam, V., Schmitz, M. H. A., Chapuis, C., Gerlich, D. W., Schneider, R., Eils, R., Huber, W., Peters, J.-M., Hyman, A. A., Durbin, R., Pepperkok, R., & Ellenberg, J. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464(7289), 721–727.
URL <http://www.nature.com/articles/nature08869>
- Nishikawa, M., Kanno, H., Zhou, Y., Xiao, T.-H., Suzuki, T., Ibayashi, Y., Harmon, J., Takizawa, S., Hiramatsu, K., Nitta, N., Kameyama, R., Peterson, W., Takiguchi, J., Shifat-E-Rabbi, M., Zhuang, Y., Yin, X., Rubaiyat, A. H. M., Deng, Y., Zhang, H., Miyata, S., Rohde, G. K., Iwasaki, W., Yatomi, Y., & Goda, K. (2021). Massive image-based single-cell profiling reveals high levels of circulating platelet aggregates in patients with COVID-19. *Nature Communications*, 12(1), 7135.
URL <https://www.nature.com/articles/s41467-021-27378-2>
- Nogales, C., Mamdouh, Z. M., List, M., Kiel, C., Casas, A. I., & Schmidt, H. H. (2021). Network pharmacology: curing causal mechanisms instead of treating symptoms. *Trends in Pharmacological Sciences*.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0165614721002212>
- Núñez-Carpintero, I., Petrizzelli, M., Zinovyev, A., Cirillo, D., & Valencia, A. (2021). The multilayer community structure of medulloblastoma. *iScience*, 24(4), 102365.
URL <https://linkinghub.elsevier.com/retrieve/pii/S2589004221003333>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., Caldas, G. V., Cheng, H., Chin, C.-S., Chow, W., de Lima, L. G., Dishuck, P. C., Durbin, R., Dvorkina, T., Fiddes, I. T., Formenti, G., Fulton, R. S., Fungtammasan, A., Garrison, E., Grady, P. G. S., Graves-Lindsay, T. A., Hall, I. M., Hansen, N. F., Hartley, G. A., Haukness, M., Howe, K., Hunkapiller, M. W., Jain, C., Jain, M., Jarvis, E. D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V. V., Marschall, T., McCartney, A. M., McDaniel, J., Miller, D. E., Mullikin, J. C., Myers, E. W., Olson, N. D., Paten, B., Peluso, P., Pevzner, P. A., Porubsky, D., Potapova, T., Rogaev, E. I., Rosenfeld, J. A., Salzberg, S. L., Schneider, V. A., Sedlazeck, F. J., Shafin, K., Shew, C. J., Shumate, A., Sims, Y., Smit, A. F. A., Soto, D. C., Sović, I., Storer, J. M., Streets, A., Sullivan, B. A.,

- Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B. P., Wenger, A., Wood, J. M. D., Xiao, C., Yan, S. M., Young, A. C., Zarate, S., Surti, U., McCoy, R. C., Dennis, M. Y., Alexandrov, I. A., Gerton, J. L., O\textquoterightNeill, R. J., Timp, W., Zook, J. M., Schatz, M. C., Eichler, E. E., Miga, K. H., & Phillippy, A. M. (2021). The complete sequence of a human genome. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/05/27/2021.05.26.445798>
- Osińska, I., Popko, K., & Demkow, U. (2014). Perforin: an important player in immune response. *Central European Journal of Immunology*, 1, 109–115.
URL <http://www.termedia.pl/doi/10.5114/ceji.2014.42135>
- Ouyang, W., Mueller, F., Hjelmare, M., Lundberg, E., & Zimmer, C. (2019a). ImJoy: an open-source computational platform for the deep learning era. *Nature Methods*, 16(12), 1199–1200.
URL <http://www.nature.com/articles/s41592-019-0627-0>
- Ouyang, W., Winsnes, C. F., Hjelmare, M., Cesnik, A. J., Åkesson, L., Xu, H., Sullivan, D. P., Dai, S., Lan, J., Jinmo, P., Galib, S. M., Henkel, C., Hwang, K., Poplavskiy, D., Tunguz, B., Wolfinger, R. D., Gu, Y., Li, C., Xie, J., Buslov, D., Fironov, S., Kiselev, A., Panchenko, D., Cao, X., Wei, R., Wu, Y., Zhu, X., Tseng, K.-L., Gao, Z., Ju, C., Yi, X., Zheng, H., Kappel, C., & Lundberg, E. (2019b). Analysis of the Human Protein Atlas Image Classification competition. *Nature Methods*, 16(12), 1254–1261.
URL <http://www.nature.com/articles/s41592-019-0658-6>
- Pang, M., Su, K., & Li, M. (2021). Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/11/28/2021.11.28.470212>
- Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7), 979–981.
URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq046>
- Peng, T., Thorn, K., Schroeder, T., Wang, L., Theis, F. J., Marr, C., & Navab, N. (2017). A BaSiC tool for background and shading correction of optical microscopy images. *Nature Communications*, 8(1), 14836.
URL <http://www.nature.com/articles/ncomms14836>
- Pereira, D. A., & Williams, J. A. (2007). Origin and evolution of high throughput screening. *British Journal of Pharmacology*, 152(1), 53–61.
URL <http://doi.wiley.com/10.1038/sj.bjp.0707373>
- Pfajfer, L., Mair, N. K., Jiménez-Heredia, R., Genel, F., Gulez, N., Ardeniz, O., Hoeger, B., Bal, S. K., Madritsch, C., Kalinichenko, A., Chandra Ardy, R., Gerçeker, B., Rey-Barroso, J., Ijspeert, H., Tangye, S. G., Simonitsch-Klupp, I., Huppa, J. B., van der Burg, M., Dupré, L., & Boztug, K. (2018). Mutations affecting the actin regulator WD repeat-containing protein 1 lead to aberrant lymphoid immunity. *The Journal of allergy and clinical immunology*, 142(5), 1589–1604.
URL <http://www.ncbi.nlm.nih.gov/pubmed/29751004>
- Piñero, J., Berenstein, A., Gonzalez-Perez, A., Chernomoretz, A., & Furlong, L. I. (2016). Uncovering disease mechanisms through network biology in the era of Next Generation

- Sequencing. *Scientific Reports*, 6(1), 24570.
URL <http://www.nature.com/articles/srep24570>
- Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., & Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, 2015, bav028–bav028.
URL <https://academic.oup.com/database/article-lookup/doi/10.1093/database/bav028>
- Pingoud, A., Wilson, G. G., & Wende, W. (2014). Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Research*, 42(12), 7489–7527.
URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku447>
- Piotrowski, J. S., Li, S. C., Deshpande, R., Simpkins, S. W., Nelson, J., Yashiroda, Y., Barber, J. M., Safizadeh, H., Wilson, E., Okada, H., Gebre, A. A., Kubo, K., Torres, N. P., LeBlanc, M. A., Andrusiak, K., Okamoto, R., Yoshimura, M., DeRango-Adem, E., van Leeuwen, J., Shirahige, K., Baryshnikova, A., Brown, G. W., Hirano, H., Costanzo, M., Andrews, B., Ohya, Y., Osada, H., Yoshida, M., Myers, C. L., & Boone, C. (2017). Functional annotation of chemical libraries across diverse biological processes. *Nature Chemical Biology*, 13(9), 982–993.
URL <http://www.nature.com/doifinder/10.1038/nchembio.2436>
- Prewitt, J. M. S., & Mendelsohn, M. L. (1966). THE ANALYSIS OF CELL IMAGES*. *Annals of the New York Academy of Sciences*, 128(3), 1035–1053.
URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1749-6632.1965.tb11715.x>
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, 152(5), 1173–1183.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867413002110>
- Qin, Y., Huttlin, E. L., Winsnes, C. F., Gosztyla, M. L., Wacheul, L., Kelly, M. R., Blue, S. M., Zheng, F., Chen, M., Schaffer, L. V., Licon, K., Bäckström, A., Vaites, L. P., Lee, J. J., Ouyang, W., Liu, S. N., Zhang, T., Silva, E., Park, J., Pitea, A., Kreisberg, J. F., Gygi, S. P., Ma, J., Harper, J. W., Yeo, G. W., Lafontaine, D. L. J., Lundberg, E., & Ideker, T. (2021). A multi-scale map of cell structure fusing protein images and interactions. *Nature*.
URL <https://www.nature.com/articles/s41586-021-04115-9>
- Regan-Fendt, K. E., Xu, J., DiVincenzo, M., Duggan, M. C., Shakya, R., Na, R., Carson, W. E., Payne, P. R. O., & Li, F. (2019). Synergy from gene expression and network mining (SynGeNet) method predicts synergistic drug combinations for diverse melanoma genomic subtypes. *npj Systems Biology and Applications*, 5(1), 6.
URL <http://www.nature.com/articles/s41540-019-0085-4>
- Reicher, A., Koren, A., & Kubicek, S. (2020). Pooled protein tagging, cellular imaging, and in situ sequencing for monitoring drug action in real time. *Genome Research*, 30(12), 1846–1855.
URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.261503.120>
- Rendeiro, A. F., Ravichandran, H., Bram, Y., Chandar, V., Kim, J., Meydan, C., Park, J., Foox, J., Hether, T., Warren, S., Kim, Y., Reeves, J., Salvatore, S., Mason, C. E., Swanson, E. C., Borczuk, A. C., Elemento, O., & Schwartz, R. E. (2021). The spatial landscape of lung

- pathology during COVID-19 progression. *Nature*, 593(7860), 564–569.
URL <http://www.nature.com/articles/s41586-021-03475-6>
- Richards, A. L., Eckhardt, M., & Krogan, N. J. (2021). Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Molecular Systems Biology*, 17(1).
URL <https://onlinelibrary.wiley.com/doi/10.15252/msb.20188792>
- Rieckmann, J. C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S. S., Lanzavecchia, A., Mann, M., & Meissner, F. (2017). Social network architecture of human immune cells unveiled by quantitative proteomics. *Nature Immunology*, 18(5), 583–593.
URL <http://www.nature.com/articles/ni.3693>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47.
URL <http://academic.oup.com/nar/article/43/7/e47/2414268/limma-powers-differential-expression-analyses-for>
- Roesch, E., Greener, J. G., MacLean, A. L., Nassar, H., Rackauckas, C., Holy, T. E., & Stumpf, M. P. H. (2021). Julia for Biologists. *arXiv*.
URL <http://arxiv.org/abs/2109.09973>
- Rohban, M. H., Abbasi, H. S., Singh, S., & Carpenter, A. E. (2019). Capturing single-cell heterogeneity via data fusion improves image-based profiling. *Nature Communications*, 10(1), 2082.
URL <http://www.nature.com/articles/s41467-019-10154-8>
- Sailem, H., Bousgouni, V., Cooper, S., & Bakal, C. (2014). Cross-talk between Rho and Rac GTPases drives deterministic exploration of cellular shape space and morphological heterogeneity. *Open Biology*, 4(1), 130132.
URL <https://royalsocietypublishing.org/doi/10.1098/rsob.130132>
- Sailem, H. Z., Sero, J. E., & Bakal, C. (2015). Visualizing cellular imaging data using PhenoPlot. *Nature Communications*, 6(1), 5825.
URL <http://www.nature.com/articles/ncomms6825>
- Salzer, E., Cagdas, D., Hons, M., Mace, E. M., Garncarz, W., Petronczki, Y., Platzer, R., Pfajfer, L., Bilic, I., Ban, S. A., Willmann, K. L., Mukherjee, M., Supper, V., Hsu, H. T., Banerjee, P. P., Sinha, P., McClanahan, F., Zlabinger, G. J., Pickl, W. F., Gribben, J. G., Stockinger, H., Bennett, K. L., Huppa, J. B., Dupré, L., Sanal, J., Jäger, U., Sixt, M., Tezcan, I., Orange, J. S., & Boztug, K. (2016). RASGRP1 deficiency causes immunodeficiency with impaired cytoskeletal dynamics. *Nature Immunology*, 17(12), 1352–1360.
URL <http://www.nature.com/articles/ni.3575>
- Salzer, E., Zoghi, S., Kiss, M. G., Kage, F., Rashkova, C., Stahnke, S., Haimel, M., Platzer, R., Caldera, M., Ardy, R. C., Hoeger, B., Block, J., Medgyesi, D., Sin, C., Shahkarami, S., Kain, R., Ziaee, V., Hammerl, P., Bock, C., Menche, J., Dupré, L., Huppa, J. B., Sixt, M., Lomakin, A., Rottner, K., Binder, C. J., Stradal, T. E. B., Rezaei, N., & Boztug, K. (2020). The cytoskeletal regulator HEM1 governs B cell development and prevents autoimmunity. *Science Immunology*, 5(49).
URL <https://www.science.org/doi/10.1126/sciimmunol.abc3979>

- Santos, A., Colaço, A. R., Nielsen, A. B., Niu, L., Geyer, P. E., Coscia, F., Albrechtsen, N. J. W., Mundt, F., Jensen, L. J., & Mann, M. (2020). Clinical Knowledge Graph Integrates Proteomics Data into Clinical Decision-Making. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2020/05/10/2020.05.09.084897>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682.
URL <http://www.nature.com/articles/nmeth.2019>
- Schmidt, H. (2018). REPO-TRIAL: Common mechanism-based drug repurposing and endophenotyping. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (pp. 1–2). IEEE.
URL <https://ieeexplore.ieee.org/document/8621156/>
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675.
URL <http://www.nature.com/articles/nmeth.2089>
- Scholes, N. S., Mayor-Ruiz, C., & Winter, G. E. (2021). Identification and selectivity profiling of small-molecule degraders via multi-omics approaches. *Cell Chemical Biology*, 28(7), 1048–1060.
URL <https://linkinghub.elsevier.com/retrieve/pii/S245194562100146X>
- Schraivogel, D., Kuhn, T. M., Rauscher, B., Rodríguez-Martínez, M., Paulsen, M., Owsley, K., Middlebrook, A., Tischer, C., Ramasz, B., Ordoñez-Rueda, D., Dees, M., Cuylen-Haering, S., Diebold, E., & Steinmetz, L. M. (2022). High-speed fluorescence image-enabled cell sorting. *Science*, 375(6578), 315–320.
URL <https://www.science.org/doi/10.1126/science.abj3013>
- Schulz, D., Zanutelli, V. R. T., Fischer, J. R., Schapiro, D., Engler, S., Lun, X.-K., Jackson, H. W., & Bodenmiller, B. (2018). Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Systems*, 6(1), 25–36.
URL <https://linkinghub.elsevier.com/retrieve/pii/S2405471217305434>
- Schuster, P. (2000). Taming combinatorial explosion. *Proceedings of the National Academy of Sciences*, 97(14), 7678–7680.
URL <http://www.pnas.org/cgi/doi/10.1073/pnas.150237097>
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, 14(8), e1002533.
URL <https://dx.plos.org/10.1371/journal.pbio.1002533>
- Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1), 146.
URL <http://electronicimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.1631315>
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., & Zhang, F. (2014). Genome-Scale CRISPR-Cas9

- Knockout Screening in Human Cells. *Science*, 343(6166), 84–87.
URL <https://www.science.org/doi/10.1126/science.1247005>
- Sharma, K., Choi, S. Y., Zhang, Y., Nieland, T. J. F., Long, S., Li, M., & Huganir, R. L. (2013). High-throughput genetic screen for synaptogenic factors: Identification of lrp6 as critical for excitatory synapse development. *Cell Reports*, 5(5), 1330–1341.
- Sharma, P. P., Bansal, M., Sethi, A., Poonam, Pena, L., Goel, V. K., Grishina, M., Chaturvedi, S., Kumar, D., & Rathi, B. (2021). Computational methods directed towards drug repurposing for COVID-19: advantages and limitations. *RSC Advances*, 11(57), 36181–36198.
URL <http://xlink.rsc.org/?DOI=D1RA05320E>
- Sharma, S., & Petsalaki, E. (2018). Application of CRISPR-Cas9 Based Genome-Wide Screening Approaches to Study Cellular Signalling Mechanisms. *International Journal of Molecular Sciences*, 19(4), 933.
URL <http://www.mdpi.com/1422-0067/19/4/933>
- Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, 59(3), 223–239.
URL <https://onlinelibrary.wiley.com/doi/10.1002/jcp.1030590302>
- Sidders, B., Karlsson, A., Kitching, L., Torella, R., Karila, P., & Phelan, A. (2018). Network-Based Drug Discovery: Coupling Network Pharmacology with Phenotypic Screening for Neuronal Excitability. *Journal of Molecular Biology*.
URL <https://www.sciencedirect.com/science/article/pii/S0022283618308088?via%3Dihub>
- Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J. K., Gustin, E., Chupakhin, V., Chong, Y. T., Vialard, J., Buijnsters, P., Velter, I., Vapirev, A., Singh, S., Carpenter, A. E., Wuyts, R., Hochreiter, S., Moreau, Y., & Ceulemans, H. (2018). Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell chemical biology*, 0(0).
URL <http://www.ncbi.nlm.nih.gov/pubmed/29503208>
- Singh, S., Bray, M., Jones, T., & Carpenter, A. (2014). Pipeline for illumination correction of images for high-throughput microscopy. *Journal of Microscopy*, 256(3), 231–236.
URL <https://onlinelibrary.wiley.com/doi/10.1111/jmi.12178>
- Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., & Horvath, P. (2015). CIDRE: an illumination-correction method for optical microscopy. *Nature Methods*, 12(5), 404–406.
URL <http://www.nature.com/articles/nmeth.3323>
- Snijder, B., Vladimer, G. I., Krall, N., Miura, K., Schmolke, A.-S., Kornauth, C., Lopez de la Fuente, O., Choi, H.-S., van der Kouwe, E., Gültekin, S., Kazianka, L., Bigenzahn, J. W., Hoermann, G., Prutsch, N., Merkel, O., Ringler, A., Sabler, M., Jeryczynski, G., Mayerhoefer, M. E., Simonitsch-Klupp, I., Ocko, K., Felberbauer, F., Müllauer, L., Prager, G. W., Korkmaz, B., Kenner, L., Sperr, W. R., Kralovics, R., Gisslinger, H., Valent, P., Kubicek, S., Jäger, U., Staber, P. B., & Superti-Furga, G. (2017). Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label,

- pilot study. *The Lancet Haematology*, 4(12), e595–e606.
URL <https://linkinghub.elsevier.com/retrieve/pii/S2352302617302089>
- Sofroniew, N., Lambert, T., Evans, K., Nunez-Iglesias, J., Bokota, G., Peña-Castellanos, G., Winston, P., Yamauchi, K., Bussonnier, M., Pop, D. D., Liu, Z., ACS, Pam, alisterburt, Buckley, G., Sweet, A., Gaifas, L., Rodríguez-Guerra, J., Migas, L., Hilsenstein, V., Bragantini, J., Lee, G. R., Hector, Freeman, J., Boone, P., Lowe, A. R., Gohlke, C., Royer, L., PIERRÉ, A., & Har-Gil, H. (2021). napari/napari: 0.4.12rc2.
URL <https://doi.org/10.5281/zenodo.5587893>
- Stenström, L., Mahdessian, D., Gnann, C., Cesnik, A. J., Ouyang, W., Leonetti, M. D., Uhlén, M., Cuylen-Haering, S., Thul, P. J., & Lundberg, E. (2020). Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. *Molecular Systems Biology*, 16(8).
URL <https://onlinelibrary.wiley.com/doi/10.15252/msb.20209469>
- Stéphanou, A., Fanchon, E., Innominato, P. F., & Ballesta, A. (2018). Systems Biology, Systems Medicine, Systems Pharmacology: The What and The Why. *Acta Biotheoretica*, (pp. 1–21).
URL <http://link.springer.com/10.1007/s10441-018-9330-2>
- Strezoska, Z., Perkett, M. R., Chou, E. T., Maksimova, E., Anderson, E. M., McClelland, S., Kelley, M. L., Vermeulen, A., & Smith, A. v. B. (2017). High-content analysis screening for cell cycle regulators using arrayed synthetic crRNA libraries. *Journal of Biotechnology*, 251, 189–200.
URL <https://www.sciencedirect.com/science/article/pii/S0168165617301712>
- Stringer, C., Michaelos, M., & Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. *bioRxiv*, (p. 2020.02.02.931238).
- Sullivan, D. P., Winsnes, C. F., Åkesson, L., Hjelmare, M., Wiking, M., Schutten, R., Campbell, L., Leifsson, H., Rhodes, S., Nordgren, A., Smith, K., Revaz, B., Finnbogason, B., Szantner, A., & Lundberg, E. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nature Biotechnology*, 36(9), 820–828.
URL <http://www.nature.com/doifinder/10.1038/nbt.4225>
- Svitkina, T. (2018). The Actin Cytoskeleton and Actin-Based Motility. *Cold Spring Harbor perspectives in biology*, 10(1), a018267.
URL <http://www.ncbi.nlm.nih.gov/pubmed/29295889><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5749151>
- Swaney, D. L., Ramms, D. J., Wang, Z., Park, J., Goto, Y., Soucheray, M., Bhola, N., Kim, K., Zheng, F., Zeng, Y., McGregor, M., Herrington, K. A., O’Keefe, R., Jin, N., VanLandingham, N. K., Foussard, H., Von Dollen, J., Bouhaddou, M., Jimenez-Morales, D., Obernier, K., Kreisberg, J. F., Kim, M., Johnson, D. E., Jura, N., Grandis, J. R., Gutkind, J. S., Ideker, T., & Krogan, N. J. (2021). A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity. *Science*, 374(6563).
URL <https://www.science.org/doi/10.1126/science.abf2911>
- Tan, J., & Martin, S. E. (2016). Validation of Synthetic CRISPR Reagents as a Tool for Arrayed Functional Genomic Screening. *PLOS ONE*, 11(12), e0168968.
URL <http://dx.plos.org/10.1371/journal.pone.0168968>

- Thomas, P. D., Hill, D. P., Mi, H., Osumi-Sutherland, D., Van Auken, K., Carbon, S., Balhoff, J. P., Albou, L.-P., Good, B., Gaudet, P., Lewis, S. E., & Mungall, C. J. (2019). Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nature Genetics*, 51(10), 1429–1433.
URL <http://www.nature.com/articles/s41588-019-0500-1>
- Thorn, K. (2016). A quick guide to light microscopy in cell biology. *Molecular Biology of the Cell*, 27(2), 219–222.
URL <https://www.molbiolcell.org/doi/10.1091/mbc.e15-02-0088>
- Tuncbag, N., Gosline, S. J. C., Kedaigle, A., Soltis, A. R., Gitter, A., & Fraenkel, E. (2016). Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. *PLOS Computational Biology*, 12(4), e1004879.
URL <http://dx.plos.org/10.1371/journal.pcbi.1004879>
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59.
URL <https://www.nature.com/articles/s43586-021-00056-9>
- Vangeel, L., Chiu, W., De Jonghe, S., Maes, P., Slechten, B., Raymenants, J., André, E., Leyssen, P., Neyts, J., & Jochmans, D. (2022). Remdesivir, Molnupiravir and Nirmatrelvir remain active against SARS-CoV-2 Omicron and other variants of concern. *Antiviral Research*, 198, 105252.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0166354222000201>
- Verstraete, N., Jurman, G., Bertagnolli, G., Ghavasieh, A., Pancaldi, V., & De Domenico, M. (2020). CovMulNet19, Integrating Proteins, Diseases, Drugs, and Symptoms: A Network Medicine Approach to COVID-19. *Network and Systems Medicine*, 3(1), 130–141.
URL <https://www.liebertpub.com/doi/10.1089/nsm.2020.0011>
- Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), 583–598.
- Vizeacoumar, F. J., van Dyk, N., S.Vizeacoumar, F., Cheung, V., Li, J., Sydorsky, Y., Case, N., Li, Z., Datti, A., Nislow, C., Raught, B., Zhang, Z., Frey, B., Bloom, K., Boone, C., & Andrews, B. J. (2010). Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. *Journal of Cell Biology*, 188(1), 69–81.
URL <https://rupress.org/jcb/article/188/1/69/35659/Integrating-highthroughput-genetic-interaction>
- Vladimer, G. I., Snijder, B., Krall, N., Bigenzahn, J. W., Huber, K. V. M., Lardeau, C.-H., Sanjiv, K., Ringler, A., Berglund, U. W., Sabler, M., de la Fuente, O. L., Knöbl, P., Kubicek, S., Helleday, T., Jäger, U., & Superti-Furga, G. (2017). Global survey of the immunomodulatory potential of common drugs. *Nature Chemical Biology*, 13(6), 681–690.
URL <http://www.nature.com/articles/nchembio.2360>
- von Chamier, L., Laine, R. F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E., Lerche, M., Hernández-Pérez, S., Mattila, P. K., Karinou, E., Holden, S., Solak, A. C., Krull, A., Buchholz, T.-O., Jones, M. L., Royer, L. A., Leterrier, C., Shechtman, Y., Jug, F., Heilemann, M.,

- Jacquemet, G., & Henriques, R. (2021). Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature Communications*, 12(1), 2276.
URL <http://www.nature.com/articles/s41467-021-22518-0>
- Wählby, C., Kametsky, L., Liu, Z. H., Riklin-Raviv, T., Conery, A. L., O'Rourke, E. J., Sokolnicki, K. L., Visvikis, O., Ljosa, V., Irazoqui, J. E., Golland, P., Ruvkun, G., Ausubel, F. M., & Carpenter, A. E. (2012). An image analysis toolbox for high-throughput *C. elegans* assays. *Nature Methods*, 9(7), 714–716.
URL <http://www.nature.com/articles/nmeth.1984>
- Walsh, B., Mohamed, S. K., & Nováček, V. (2020). BioKG: A Knowledge Graph for Relational Learning On Biological Data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, (pp. 3173–3180). New York, NY, USA: ACM.
URL <https://dl.acm.org/doi/10.1145/3340531.3412776>
- Wang, C., Lu, T., Emanuel, G., Babcock, H. P., & Zhuang, X. (2019). Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proceedings of the National Academy of Sciences*, 116(22), 10842–10851.
URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1903808116>
- Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*, 343(6166), 80–84.
URL <https://www.science.org/doi/10.1126/science.1246981>
- Wassie, A. T., Zhao, Y., & Boyden, E. S. (2019). Expansion microscopy: principles and uses in biological research. *Nature Methods*, 16(1), 33–41.
URL <http://www.nature.com/articles/s41592-018-0219-4>
- Wawer, M. J., Li, K., Gustafsdottir, S. M., Ljosa, V., Bodycombe, N. E., Marton, M. A., Sokolnicki, K. L., Bray, M.-A., Kemp, M. M., Winchester, E., Taylor, B., Grant, G. B., Hon, C. S.-Y., Duvall, J. R., Wilson, J. A., Bittker, J. A., Danik, V., Narayan, R., Subramanian, A., Winckler, W., Golub, T. R., Carpenter, A. E., Shamji, A. F., Schreiber, S. L., & Clemons, P. A. (2014). Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proceedings of the National Academy of Sciences*, 111(30), 10911–10916.
URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1410933111>
- Way, G. P., Kost-Alimova, M., Shibue, T., Harrington, W. F., Gill, S., Piccioni, F., Becker, T., Shafqat-Abbasi, H., Hahn, W. C., Carpenter, A. E., Vazquez, F., & Singh, S. (2021a). Predicting cell health phenotypes using image-based morphology profiling. *Molecular Biology of the Cell*, (pp. 20–12).
URL <https://www.molbiolcell.org/doi/10.1091/mbc.E20-12-0784>
- Way, G. P., Natoli, T., Adeboye, A., Litichevskiy, L., Yang, A., Lu, X., Caicedo, J. C., Cimini, B. A., Karhohs, K., Logan, D. J., Rohban, M., Kost-Alimova, M., Hartland, K., Bornholdt, M., Chandrasekaran, N., Haghighi, M., Singh, S., Subramanian, A., & Carpenter, A. E. (2021b). Morphology and gene expression profiling provide complementary information for mapping cell state. *bioRxiv*.
URL <https://www.biorxiv.org/content/early/2021/10/22/2021.10.21.465335>
- Weinstein, J. A., Regev, A., & Zhang, F. (2019). DNA Microscopy: Optics-free Spatio-genetic

- Imaging by a Stand-Alone Chemical Reaction. *Cell*, 178(1), 229–241.
URL <http://www.ncbi.nlm.nih.gov/pubmed/31230717>
- Westhoff, J. H., Steenbergen, P. J., Thomas, L. S. V., Heigwer, J., Bruckner, T., Cooper, L., Tönshoff, B., Hoffmann, G. F., & Gehrig, J. (2020). In vivo High-Content Screening in Zebrafish for Developmental Nephrotoxicity of Approved Drugs. *Frontiers in Cell and Developmental Biology*, 8.
URL <https://www.frontiersin.org/article/10.3389/fcell.2020.00583/full>
- Wheeler, E. C., Vu, A. Q., Einstein, J. M., DiSalvo, M., Ahmed, N., Van Nostrand, E. L., Shishkin, A. A., Jin, W., Allbritton, N. L., & Yeo, G. W. (2020). Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. *Nature Methods*.
URL <http://www.nature.com/articles/s41592-020-0826-8>
- White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The Structure of the Nervous System of the Nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 314(1165), 1–340.
URL <http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.1986.0056>
- Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R. K., Sarkans, U., Brazma, A., Carazo Salas, R. E., & Swedlow, J. R. (2017). Image Data Resource: a bioimage data integration and publication platform. *Nature Methods*, 14(8), 775–781.
URL <http://www.nature.com/articles/nmeth.4326>
- Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell*, 173(7), 1573–1580.
URL <http://www.ncbi.nlm.nih.gov/pubmed/29906445>
- Wright, A. V., Nuñez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering. *Cell*, 164(1-2), 29–44.
URL <http://www.ncbi.nlm.nih.gov/pubmed/26771484>
- Yeger-Lotem, E., & Sharan, R. (2015). Human protein interaction networks across tissues and diseases. *Frontiers in Genetics*, 6.
URL <http://journal.frontiersin.org/Article/10.3389/fgene.2015.00257/abstract>
- Zamboni, V., Jones, R., Umbach, A., Ammoni, A., Passafaro, M., Hirsch, E., Merlo, G., Zamboni, V., Jones, R., Umbach, A., Ammoni, A., Passafaro, M., Hirsch, E., & Merlo, G. R. (2018). Rho GTPases in Intellectual Disability: From Genetics to Therapeutic Opportunities. *International Journal of Molecular Sciences*, 19(6), 1821.
URL <http://www.mdpi.com/1422-0067/19/6/1821>
- Zewail, A. H. (2010). Four-Dimensional Electron Microscopy. *Science*, 328(5975), 187–193.
URL <https://www.science.org/doi/10.1126/science.1166135>
- Zhang, W., Chien, J., Yong, J., & Kuang, R. (2017). Network-based machine learning and graph theory algorithms for precision oncology. *npj Precision Oncology*, 1(1), 25.
URL <http://www.nature.com/articles/s41698-017-0029-7>
- Zheng, F., Kelly, M. R., Ramms, D. J., Heintschel, M. L., Tao, K., Tutuncuoglu, B., Lee, J. J., Ono, K., Foussard, H., Chen, M., Herrington, K. A., Silva, E., Liu, S. N., Chen, J., Churas, C., Wilson, N., Kratz, A., Pillich, R. T., Patel, D. N., Park, J., Kuenzi, B., Yu, M. K., Licon,

K., Pratt, D., Kreisberg, J. F., Kim, M., Swaney, D. L., Nan, X., Fraley, S. I., Gutkind, J. S., Krogan, N. J., & Ideker, T. (2021). Interpretation of cancer mutations using a multiscale map of protein systems. *Science*, 374(6563).

URL <https://www.science.org/doi/10.1126/science.abf3067>

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell discovery*, 6(1), 14.

URL <http://www.ncbi.nlm.nih.gov/pubmed/32194980><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7073332>

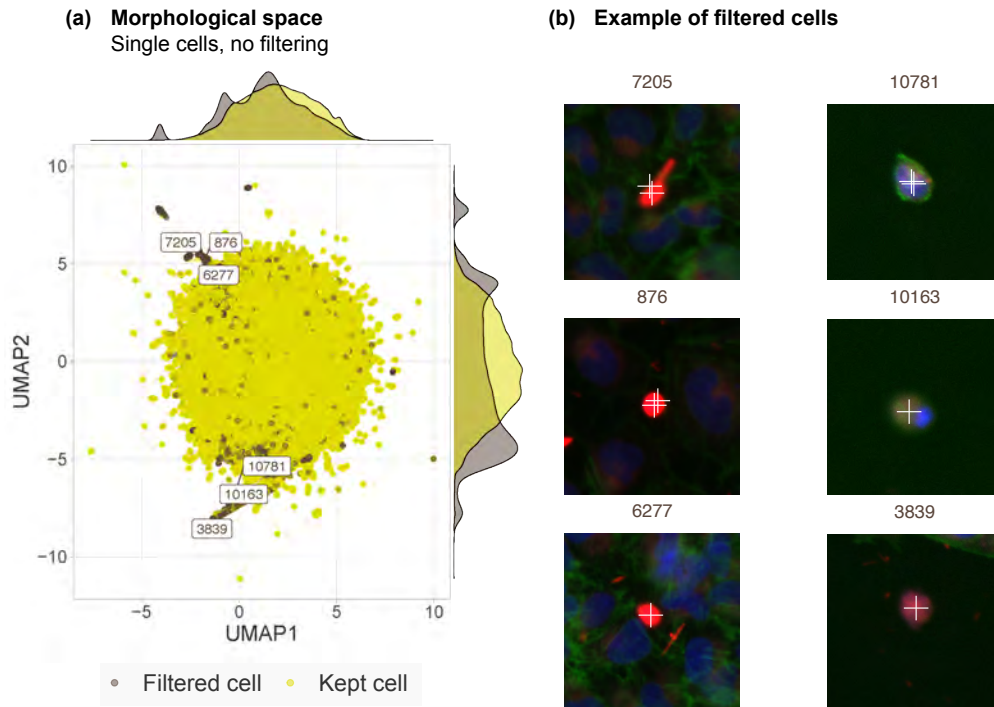
Ziegler, S., Sievers, S., & Waldmann, H. (2021). Morphological profiling of small molecules. *Cell Chemical Biology*, 28(3), 300–319.

URL <https://linkinghub.elsevier.com/retrieve/pii/S2451945621000970>

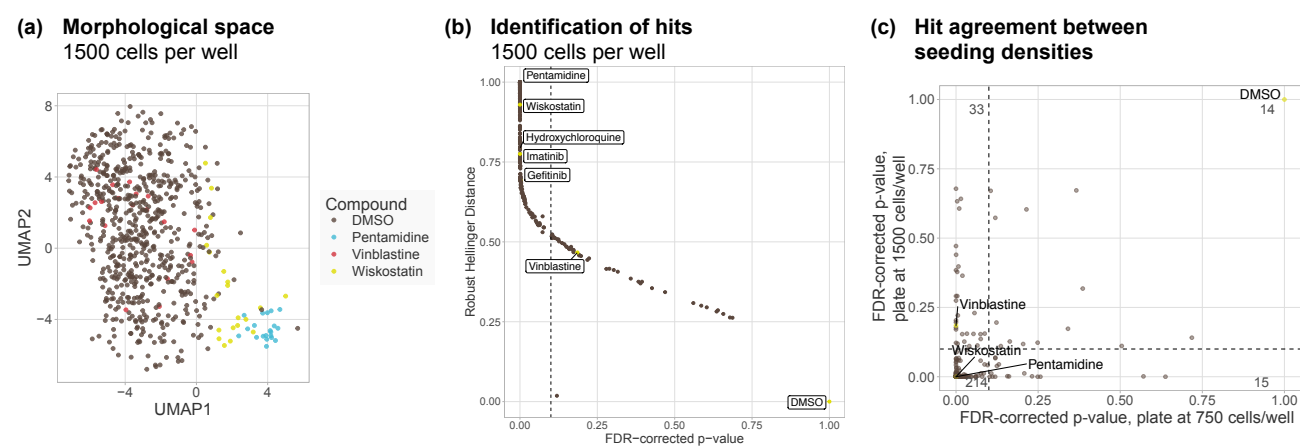
Appendix

Supplementary material - BioProfiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations

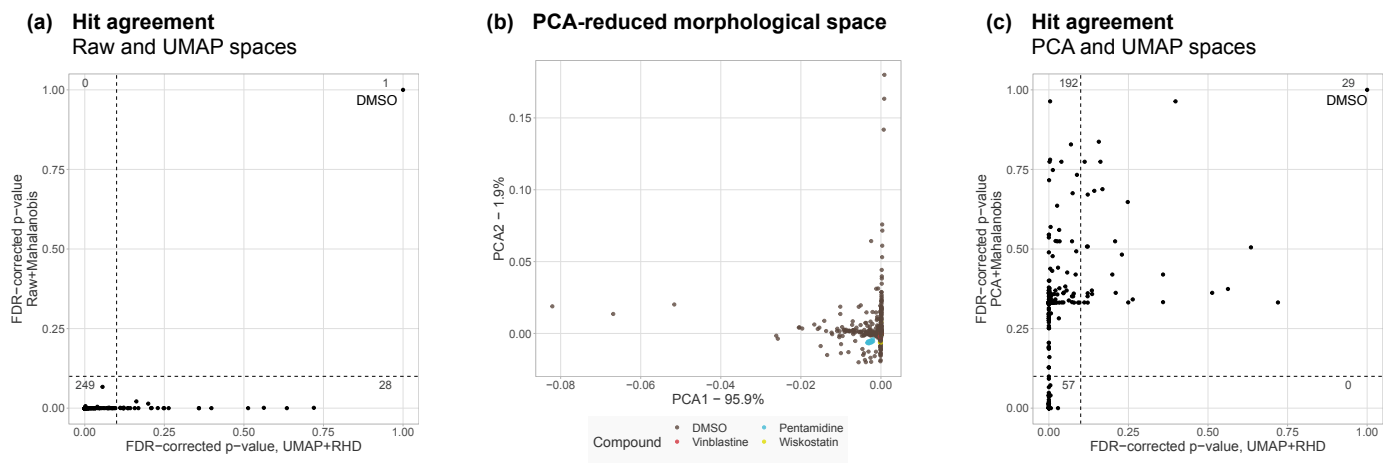
This section reproduces the supplementary material corresponding to the paper entitled “Bio-Profiling.jl: Profiling biological perturbations with high-content imaging in single cells and heterogeneous populations” presentend in section 3.1.



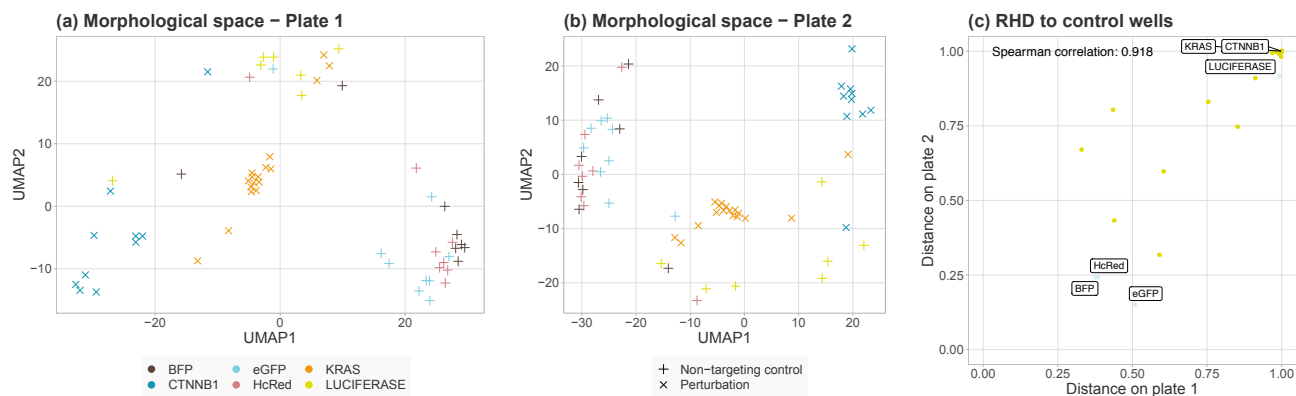
Supp. Figure 1: Filtering of artifacts and biological outliers is required for defining an informative morphological space. **(a)** UMAP embedding preserving the Euclidean distance between the morphological profiles of single cells in the plate seeded with 750 cells per well and without any filtering. Cells that we excluded from further analysis based on a set of feature thresholds (see Methods) are highlighted in brown. Six individual cells are labeled with a unique identifier. Marginal plots show relative density of filtered and kept cells. **(b)** Cropped images of six example cells discarded from further analyses along with their unique identifier. Channels are CellMask (red), phalloidin (green) and DAPI (blue). All images are shown at the same scale.



Supp. Figure 2: BioProfiling.jl profiles of plates seeded with 750 and 1500 cells per well show similarities. (a) UMAP embedding preserving the cosine distance between the morphological profiles aggregated per field of view in the plate seeded with 1500 cells per well. Two out of four dimensions are represented. **(b)** Robust Hellinger distance and Robust Morphological Perturbation Value (FDR-corrected p-value) of each compound in the plate seeded with 1500 cells per well compared to DMSO. Vertical dotted line indicates an FDR threshold of 0.1 and all compounds on its left are defined as morphological hits. **(c)** FDR-corrected p-value of the significance of morphological changes induced by each compound in both plates. Dotted lines indicate an FDR threshold of 0.1 and a number indicates the number of points in each of the regions delineated by the dotted lines.



Supp. Figure 3: BioProfiling.jl supports various approaches to hit detection. (a) FDR-corrected p-value of the significance of morphological changes induced by each compound based either on the robust Hellinger distance in the UMAP-reduced space or the Mahalanobis distance to the center of the DMSO profiles in the original space. (b) PCA embedding of the morphological profiles aggregated per field of view in the plate seeded with 750 cells per well. (c) FDR-corrected p-value of the significance of morphological changes induced by each compound based either on the robust Hellinger distance in the UMAP-reduced space or the Mahalanobis distance to the center of the DMSO profiles in the PCA space. In (a) and (c), dotted lines indicate an FDR threshold of 0.1 and a number indicates the number of points in each of the regions delineated by the dotted lines.



Supp. Figure 4: BioProfiling.jl generates robust morphological profiles from CellPainting experiments. (a-b) UMAP embedding preserving the Euclidean distance between the morphological profiles aggregated per well in plate 1 (a) or plate 2 (b), subsetting to non-targeting controls and overexpression of CTNNB1 and KRAS. **(c)** Robust Hellinger Distance (RHD) between each perturbation and the non-targeting profiles in both plates. Data source: Cell Painting Image Collection (see Methods).

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
(+)-Butaclamol hydrochloride			0.2479	0	0.02
(+)-Cyclazocine			0.0288	0.0012	20
(+/-)-Sulfinpyrazone	["Uricosuric blocker"]	["ABCC1", "ABCC2", "FPR1", "SLC22A12"]	0.0019	0.0154	20
(-)-JQ1			0.0004	0	20
(-)-Perillic acid			0.157	0.0018	20
(-)-Quinpirole hydrochloride	["Dopamine receptor agonist"]	["DRD2", "DRD3", "DRD4", "DRD1", "HTR1A", "HTR2A", "HTR2B", "HTR2C"]	0.0016	0.1061	0.2
(-)-trans-(1S,2S)-U-50488 hydrochloride			0.0087	0.0278	20
(S)-(+)-Camptothecin				0	20
(S)-Propranolol hydrochloride	["Adrenergic receptor antagonist"]	["ADRB2", "ADRB3", "ADRB1", "CYP2C19", "HTR1A", "HTR1B"]	0.0287	0	2
(±)-Isoproterenol hydrochloride					20
(±)-Methoxyverapamil hydrochloride					0.2
(±)-Metoprolol (+)-tartrate					0.2
(±)-Octoclothebin maleate					20
(±)-SKF-38393 hydrochloride					20
(±)-Sulpiride					0.02
(±)-Verapamil hydrochloride					2
(±)-alpha-Lipoic Acid					2
1,10-Phenanthroline monohydrate			0.0208	0.0026	2
1,7-Dimethylxanthine			0.0558	0.1604	20
2,2'-Bipyridyl			0.0013	0.0308	20
2,3-Dimethoxy-1,4-naphthoquinone			0.0055	0.0642	0.02
2-Phenylaminoadenosine			0.1222	0	20
2-methoxyestradiol			0	0.1733	2
4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride			0	0.0015	20
4-Hydroxy-3-methoxyphenylacetic acid			0	0.1915	20
5-(N,N-hexamethylene)amiloride			0	0	20.000633
5-(N-Ethyl-N-isopropyl)amiloride			0	0	19.999068
5-Bromo-2'-deoxyuridine			0	0	20
5-Fluorouracil			0	0	20
5-azacytidine			0	0	19.99822
5HPP-33			0.398	0.3161	0.02
5alpha-Pregnan-3alpha-ol-20-one			0.0173	0	20
6,7-ADTN hydrobromide			0	0.3031	0.02
6-Nitroso-1,2-benzopyrone			0	0	0.02
7-Cyclopentyl-5-(4-phenoxy)phenyl-7H-pyrrolo[2,3-d]pyrimidin-4-ylamine			0.0016	0	20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
A-77636 hydrochloride					20.000382
AC-93253 iodide			0	0	0.2
AEG 3482			0	0	20
AMG 9810	["TRPV antagonist"]	["TRPV1"]	0.0314	0.0003	2
AZ191					20
Acepromazine maleate	["Dopamine receptor antagonist"]	["ADRA1A", "ADRA1B", "DRD1", "DRD2", "HTR1A", "HTR2A"]	0.1114	0.0097	0.02
Adaphostin					20
Albendazole	["Anthelmintic", "Tubulin inhibitor"]	["CYP1A2", "CYP2J2", "TUBA1A", "TUBB", "TUBB4B"]	0	0	20
Aliskiren	["Antihypertensive", "Peptidase inhibitor", "Protease inhibitor", "Renin inhibitor"]	["REN"]	0.0063	0.2234	20
Ammonium pyrrolidinedithiocarbamate				0	20
Amodiaquine	["Histamine receptor agonist"]	["HNMT", "CYP2C8"]	0	0	20
Amsacrine hydrochloride	["Topoisomerase inhibitor"]	["TOP2A", "KCNH2"]	0.2101	0.5859	2
Ancitabine hydrochloride			0	0	0.2
Apomorphine hydrochloride hemihydrate		["DRD2", "DRD5", "ADRA2A", "ADRA2B", "ADRA2C", "DRD1", "DRD3", "DRD4", "HTR1A", "HTR2A", "HTR2B", "HTR2C", "TRPA1", "CALY", "HTR1B", "HTR1D"]	0.0126	0.0012	1.9999762
Arbidol hydrochloride			0.0039	0.1244	2
Auranofin					20
Aurora-A Inhibitor I			0.1221	0.0731	20.000985
Aurothioglucose			0		20
Azatadine			0.1125	0.6573	20
Azithromycin	["Bacterial 50S ribosomal subunit inhibitor"]	["MLNR"]	0.1359	0.1024	2
BAY 61-3606 hydrochloride hydrate				0	20
BIO			0	0.0003	20
BIX 01294 trihydrochloride hydrate					20
BMS-193885				0	20
BRD3308			0	0	20
BTO-1			0.0842	0.0118	19.998522
BW 723C86	["Serotonin receptor agonist"]	["HTR2B", "HTR2A", "HTR2C"]	0.0126	0	0.02
Bay 11-7082					20
Bay 11-7085					19.999076
Benidipine hydrochloride	["Calcium channel blocker"]	["CACNA1C", "CACNA1G", "CYP3A5"]	0	0	20
Benoxathian hydrochloride			0.0004	0	20
Benzamil hydrochloride		["PKD2L1", "SCNN1A", "SCNN1B", "SCNN1G", "ASIC1", "SCNN1D", "SLC8A1"]	0.0004	0	20.000339
Benztropine mesylate			0	0	20.00068
Bortezomib					20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Brefeldin A from Penicillium brefeldianum	["Protein synthesis inhibitor", "Brefeldin A inhibited guanine nucleotide exchange protein inhibitor", "Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor inhibitor"]	["ARF1", "ARFGEF1", "ARFGEF2", "CYTH2", "GBF1", "SAR1A"]	0	0	20
Brequinar sodium salt hydrate			0	0	0.2
Budesonide		["NR3C1", "CYP3A5", "CYP3A7"]	0	0	19.999503
CCCI-01			0	0	20
CCT137690				0	2
CGS-15943	["Adenosine receptor antagonist"]	["ADORA1", "ADORA2A", "ADORA2B", "ADORA3"]	0.1351	0.5604	2
CID 11210285 hydrochloride			0	0	20
CID2858522			0.0053	0	20
CP466722	["ATM kinase inhibitor"]	["ATM"]	0	0	20
CYM50358			0	0	0.2
Caffeic Acid	["Lipoxygenase inhibitor", "HIV integrase inhibitor", "NFkB pathway inhibitor", "Nitric oxide production inhibitor", "PPAR receptor modulator", "TNF production inhibitor", "Tumor necrosis factor production inhibitor"]	["ALOX5", "MIF", "RELA", "TNF"]	0	0.0062	0.0200001
Caffeic acid phenethyl ester		["RELA"]	0.0326	0	0.02
Calcimycin					20
Cantharidic Acid			0.0071	0.0097	20
Cantharidin		String[]	0.0053	0	20
Carmofur	["Thymidylate synthase inhibitor"]	["TYMS"]	0	0	20
Carvedilol	["Adrenergic receptor antagonist"]	["ADRB1", "ADRB2", "ADRA1A", "ADRA1B", "ADRA1D", "ADRA2A", "ADRA2B", "ADRA2C", "ADRB3", "CYP2C19", "CYP2E1", "GJA1", "HIF1A", "KCNH2", "NDUFC2", "NPPB", "RYSR2", "SELE", "VCAM1", "VEGFA"]	0	0	20
Cerivastatin					20
Chloroquine	["Antimalarial"]	["CYP2C8", "GSTA2", "MRGPRX1", "TLR9", "TNF"]	0	0	20
Chlorpromazine hydrochloride	["Dopamine receptor antagonist"]	["DRD2", "ADRA2A", "ADRA2B", "ADRA2C", "DRD1", "DRD3", "DRD4", "DRD5", "HRH1", "HTR1A", "HTR2A", "HTR2C", "HTR6", "HTR7", "ADRA1A", "ADRA1B", "ADRA1D", "CALM1", "CHRM1", "CHRM3", "HRH4", "HTR2B", "KCNH2", "KIF11", "ORM1", "ORM2", "SMPD1", "TRPC5"]	0.0022	0.6397	0.02

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Chlorprothixene hydrochloride	["Dopamine receptor antagonist"]	["DRD2", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "DRD1", "DRD3", "HRH1", "HTR2A", "HTR2B", "HTR2C"]	0	0	20
Cilengitide trifluoroacetic acid salt			0	0	20
Cilnidipine	["Calcium channel blocker"]	["CACNA1B", "CACNA1C"]	0.0007	0	20
Cinacalcet					20
Clemizole hydrochloride		String[]	0.2636	0	21
Clodronic acid		["SLC25A4", "SLC25A5", "SLC25A6"]	0.001	0.0635	20
Clofarabine	["Ribonucleoside reductase inhibitor"]	["RRM1", "POLA1", "RRM2", "SLC22A8"]	0.0332	0	2
Clomipramine hydrochloride	["Serotonin transporter inhibitor (SERT)"]	["SLC6A4", "SLC6A2", "CYP2C19", "GSTP1", "HTR2A", "HTR2B", "HTR2C", "SLC6A3"]	0.1222	0.041	2
Clotrimazole	["Cytochrome P450 inhibitor", "Imidazoline receptor ligand"]	["KCNN4", "CYP3A4", "CYP51A1", "NR1I2", "NR1I3", "TRPM2", "TRPM4", "TRPM8"]	0	0	20
Colchicine		["TUBB", "GLRA1", "GLRA2", "TUBB1"]	0.0167	0.6355	2
Cyclosporin A		["PPIA", "ABCB11", "CAMLG", "CYP3A5", "CYP3A7", "FPR1", "PPID", "PPIF", "PPP3CA", "PPP3R2", "SLC10A1", "SLCO1B1", "SLCO1B3"]	0	0	20
Cyproterone acetate	["Androgen receptor antagonist", "Progesterone receptor agonist", "Testosterone receptor antagonist"]	["AR", "ADORA1", "ESR1"]	0	0.0178	0.02
Cytarabine	["Ribonucleotide reductase inhibitor"]	["POLB", "POLA1"]	0	0	20
Cytosine-1-beta-D-arabinofuranoside hydrochloride				0	8.86
D-Cycloserine	["Bacterial cell wall synthesis inhibitor"]	["GRIN1"]	0.0791	0.0635	35.985971
D-ribofuranosylbenzimidazole			0.0045	0.0003	20
DCEBIO	["Potassium channel activator"]	["KCNN2", "KCNN3", "KCNN4"]	0	0.6868	20
Danshensu sodium salt			0.0039	0.2793	20
Dantrolene sodium	["Calcium channel blocker"]	["RYR1", "RYR3"]	0.635	0	2
Dequalinium chloride hydrate		["KCNN1", "KCNN3"]	0	0	2
Digitoxin					20
Dihydroartemisinin			0	0	20
Dihydroergotamine		["HTR1D", "HTR1B", "ADRA2A", "DRD2", "HTR1E", "HTR1F", "HTR2B", "HTR7"]	0.1428	0.0401	20
Dihydroergotamine methanesulfonate		["HTR1D", "HTR1B", "ADRA2A", "DRD2", "HTR1E", "HTR1F", "HTR2B", "HTR7"]	0.0034	0.007	20
Dihydroouabain			0.0053	0	2
Diphenyleneiodonium chloride	["Nitric oxide synthase inhibitor"]	["ALDH1A2", "ALDH2", "ALDH5A1", "ALDH7A1", "NOX3", "XDH"]	0	0	20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Dipyridamole	["Phosphodiesterase inhibitor"]	["ADA", "PDE5A", "PDE10A", "PDE4A", "PDE7B", "PDE8A", "PDE8B", "SLC29A1"]	0.3587	0.6785	2
Docetaxel		["TUBB", "BCL2", "MAP2", "MAP4", "MAPT", "NR1I2", "TUBB1"]	0	0	2
Domperidone	["Dopamine receptor antagonist"]	["DRD2", "DRD3", "ABCG2", "CYP3A5"]	0	0	0.2
Doxazosin mesylate	["Adrenergic receptor antagonist"]	["ADRA1D", "ADRA1A", "ADRA1B", "CYP2C19", "KCNH2", "KCNH6", "KCNH7"]	0.0107	0.6036	2
Doxorubicin		["TOP2A"]	0	0	2
Doxycycline hydrochloride	["Bacterial 30S ribosomal subunit inhibitor", "Metalloproteinase inhibitor"]	["MMP8", "MMP1"]	0.0004	0	20
Droperidol	["Dopamine receptor antagonist"]	["DRD2", "ADRA1A"]	0.0045	0.3943	20
E-64			0	0.1733	20
Ebastine			0	0	20
Efavirenz	["HIV protease inhibitor"]	["CYP2B6", "CYP2C19", "CYP2C8", "CYP3A4", "CYP3A5"]	0.1994	0.014	0.0224
Ellipticine	["Topoisomerase inhibitor"]	["TOP2A", "TOP2B"]	0	0	2
Emetine dihydrochloride hydrate	["Protein synthesis inhibitor"]	["RPS2"]	0	0	20
Enclomiphene hydrochloride					20
Endoxifen				0	20
Eptifibatide			0	0.1203	20
Ethinyl Estradiol	["DNA directed DNA polymerase stimulant", "Estrogenic component in oral contraceptives", "Estrogen receptor agonist"]	["CYP2C8", "ESR1", "ESR2", "NR1I2"]	0.0958	0.0479	0.02
Etoposide		["TOP2A", "CYP2E1", "CYP3A5", "TOP2B"]	0	0	20
Flunarizine dihydrochloride	["Calcium channel blocker"]	["CACNA1G", "CACNA1H", "CACNA1I", "CALM1", "CYP2J2", "HRH1"]	0.0112	0	20
Fluoxetine hydrochloride	["Selective serotonin reuptake inhibitor (SSRI)"]	["SLC6A4", "ANO1", "CYP2C19", "HTR2B"]	0	0	20
Fluspirilene	["Dopamine receptor antagonist"]	["DRD2", "HTR2A", "CACNG1", "HRH1", "HTR1A", "HTR1D", "HTR1E"]	0.0206	0	20
Forskolin		["ADCY2", "ADCY5", "GNAS"]	0.0013	0.0003	2
Furamidine dihydrochloride			0	0	20
GANT61			0	0.0003	20
GBR-12909 dihydrochloride			0	0	20
GSK-650394		["SGK1", "SGK2"]	0	0.0107	0.02
GSK1210151A			0	0	20
GW2974			0	0.0095	20
GW9662			0.0048	0	20
Gefitinib	["EGFR inhibitor"]	["EGFR", "CYP2C19"]	0.0403	0.0024	0.02
Gemcitabine hydrochloride	["Ribonucleotide reductase inhibitor"]	["RRM1", "CMPK1", "RRM2", "TYMS"]	0	0	20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Histamine, R(-)-alpha-methyl-, dihydrochloride			0.5132	0.1118	0.02
Hydroquinone			0.0728	0.1129	0.02
Hydroxychloroquine		["TLR7", "TLR9"]	0.0763	0	20
IKK-16 dihydrochloride	["IKK inhibitor"]	["IKKBK"]	0	0	0.02
IMS2186			0	0	20
IN-1130			0	0	20.000786
Icaritin			0	0.4267	20
Idarubicin		["TOP2A"]	0	0	20
Idazoxan hydrochloride		["NISCH"]	0	0	2
Imatinib	["BCR-ABL kinase inhibitor", "KIT inhibitor", "PDGFR receptor inhibitor"]	["ABL1", "KIT", "PDGFRA", "BCR", "CSF1R", "PDGFRB", "ABCG2", "CYP2C19", "CYP2C8", "CYP3A5", "DDR1", "NTRK1", "RET"]	0.1198	0	2
Imatinib mesylate	["BCR-ABL kinase inhibitor", "KIT inhibitor", "PDGFR receptor inhibitor"]	["ABL1", "KIT", "PDGFRA", "BCR", "CSF1R", "PDGFRB", "ABCG2", "CYP2C19", "CYP2C8", "CYP3A5", "DDR1", "NTRK1", "RET"]	0.0039	0	20
Imipramine hydrochloride	["Norepinephrine reuptake inhibitor", "Serotonin reuptake inhibitor"]	["SLC6A2", "SLC6A4", "CHRM2", "ADRA1A", "ADRA1B", "ADRA1D", "CHRM1", "CHRM3", "CHRM4", "CHRM5", "CYP2C19", "DRD1", "DRD2", "DRD5", "HRH1", "HTR1A", "HTR2A", "HTR2C", "HTR6", "HTR7", "KCND2", "KCND3", "KCNH1", "KCNH2", "SLC6A3"]	0	0	20
Iodoacetamide			0	0	20
Irinotecan					20
Isoproterenol			0.1221	0.0733	20
JFD00244			0	0	20
JS-K		String[]	0.0066	0.0012	20
K114			0	0	20
KB-R7493			0	0	20
KT203			0	0.0125	2
KU-55933	["ATM kinase inhibitor"]	["ATM", "PRKDC"]	0.0016	0.001	2
KY-05009			0	0	2.0001077
Kenpaullone	["CDK inhibitor", "Glycogen synthase kinase inhibitor"]	["GSK3B", "CDK1", "CDK5", "CCNB1", "CDK2", "LCK"]	0	0	20
Ketoconazole	["Sterol demethylase inhibitor"]	["AR", "CYP19A1", "CYP21A2", "CYP2C19", "CYP3A5", "CYP3A7", "KCN10"]	0.0013	0	20
Ketotifen fumarate	["Histamine receptor agonist", "Histamine receptor ligand", "Leukotriene receptor antagonist", "Phosphodiesterase inhibitor"]	["HRH1", "PDE4A", "PDE4B", "PDE4C", "PDE4D", "PDE7A", "PDE7B", "PDE8A", "PDE8B", "PGD"]	0.0063	0.3895	20
L-703,606 oxalate salt hydrate			0.0861	0.168	2
L-741,626			0	0	20
L-Cycloserine			0.073	0.1592	0.2

Fig. 2 Robust morphological profiling of chemical perturbations using

BioProfiling.jl.

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
L-Tryptophan			0.0007	0.0354	2
LDN-214117			0.026	0.0084	2
LP 12 hydrochloride hydrate			0.0007	0	20
LP44			0	0	20
LY-294,002 hydrochloride			0	0	20
Lasofloxifene tartrate					0.2
Lercanidipine hydrochloride hemihydrate			0.0019	0	20
Levetiracetam	["Calcium channel blocker"]	["SV2A", "CACNA1B", "SCN1A"]	0.0264	0.1362	20
Loperamide	["Opioid receptor agonist"]	["OPRM1", "OPRD1", "CACNA1A", "CALM1", "CYP2B6", "CYP2C8", "NPR2", "OPRK1", "POMC"]	0	0	20
Loperamide hydrochloride	["Opioid receptor agonist"]	["OPRM1", "OPRD1", "CACNA1A", "CALM1", "CYP2B6", "CYP2C8", "NPR2", "OPRK1", "POMC"]	0	0	20
Loratadine	["Histamine receptor antagonist"]	["HRH1", "CYP2C19", "CYP3A5"]	0.1217	0.1974	0.02
Lorcainide hydrochloride			0	0.0021	2
Lubeluzole dihydrochloride			0	0	20
M-110			0	0	20
MG 624			0	0	20
MK-677			0	0	0.2
ML-7	["Myosin light chain kinase inhibitor"]	["MYLK"]	0.0016	0	20
ML240			0.0882	0.0404	0.02
ML324			0	0	20
Maprotiline	["Norepinephrine reuptake inhibitor", "Tricyclic antidepressant"]	["SLC6A2", "ADRA1A", "ADRA1B", "ADRA1D", "ADRA2A", "ADRA2B", "ADRA2C", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "DRD2", "HRH1", "HTR2A", "HTR2C", "HTR7"]	0.0395	0.0015	2
Maprotiline hydrochloride	["Norepinephrine reuptake inhibitor", "Tricyclic antidepressant"]	["SLC6A2", "ADRA1A", "ADRA1B", "ADRA1D", "ADRA2A", "ADRA2B", "ADRA2C", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "DRD2", "HRH1", "HTR2A", "HTR2C", "HTR7"]	0	0	20
Metergoline					20
Methiothepin mesylate			0	0	20
Methoxamine hydrochloride		["ADRA1A", "ADRA1B", "ADRA1D"]	0.0013	0.0537	0.2
Metrazoline oxalate			0	0	20
Mibefradil dihydrochloride	["T-type calcium channel blocker"]	["CACNA1G", "CACNA1H", "CACNA1C", "CACNA1I", "ANO1", "CACNA1D", "CACNA1F", "CACNA1S", "CACNB1", "CACNB2", "CACNB3", "CACNB4", "CATSPER1", "CATSPER2", "CATSPER3", "CATSPER4", "CYP3A5", "CYP3A7", "SCN2A", "SCN4A", "SCN5A", "SCN9A"]	0	0	0.2

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Mifepristone	["Glucocorticoid receptor antagonist", "Progesterone receptor antagonist"]	["PGR", "NR3C1", "AR", "CYP2B6", "CYP2C8", "CYP3A5", "CYP3A7", "NR1I2"]	0.0727	0.0049	20
Mitotane	["Antineoplastic"]	["CYP11B1", "CYP11A1", "CYP3A4", "ESR1", "FDX1"]	0.0165	0.1653	20
Mitoxantrone	["Topoisomerase inhibitor"]	["TOP2A", "PIM1"]	0.7198	0.1515	2
Mycophenolic Acid	["Dehydrogenase inhibitor", "Inositol monophosphatase inhibitor"]	["IMPDH1", "IMPDH2"]	0	0	20
N-p-Tosyl-L-phenylalanine chloromethyl ketone			0.0004	0.0015	2
NG-Monomethyl-L-arginine acetate			0.0577	0	20
Nestorone			0.0007	0.0594	19.998523
Nicardipine hydrochloride	["Calcium channel blocker"]	["CACNA1C", "ADORA3", "ADRA1A", "ADRA1B", "ADRA1D", "CACNA1D", "CACNA2D1", "CACNB2", "CALM1", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "PDE1A", "PDE1B"]	0	0.0764	0.2
Niclosamide	["DNA replication inhibitor", "STAT inhibitor"]	["STAT3"]	0.0457	0	2
Nisoldipine	["Calcium channel blocker"]	["CACNA1C", "CACNA1D", "CACNA1S", "CACNA2D1", "CACNB2", "CYP3A5"]	0	0	20
Nitidine chloride					20
Nocodazole	["Tubulin inhibitor"]	["HPGDS"]	0	0	20
Nortriptyline hydrochloride	["Tricyclic antidepressant"]	["KCNJ10", "SLC6A2", "SLC6A4", "ADRA1A", "ADRA1B", "ADRA1D", "ADRA2A", "ADRA2B", "ADRA2C", "ADRB1", "ADRB2", "ADRB3", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "CYP2C19", "DRD2", "HRH1", "HTR1A", "HTR2A", "HTR2C", "HTR6", "PGRMC1", "PIK3CD", "SIGMAR1"]	0	0	20
Olanzapine	["Dopamine receptor antagonist", "Serotonin receptor antagonist"]	["DRD2", "HTR2A", "HTR2C", "DRD1", "DRD3", "DRD4", "HRH1", "HTR1A", "HTR1B", "HTR1D", "HTR1E", "HTR6", "HTR7", "ADRA1A", "ADRA1B", "ADRA2A", "ADRA2B", "ADRA2C", "ADRB1", "ADRB2", "ADRB3", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "CYP2C8", "DRD5", "GABRA1", "GABRA2", "GABRA3", "GABRA4", "GABRA5", "GABRA6", "GABRB1", "GABRB2", "GABRB3", "GABRD", "GABRE", "GABRG1", "GABRG2", "GABRG3", "GABRP", "GABRQ", "HRH2", "HRH4", "HTR1F", "HTR2B", "HTR3A", "HTR5A"]	0.0076	0.0777	20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Ouabain	["ATPase inhibitor"]	["ATP1A1", "ATP1A2", "ATP1A3", "ATP1A4", "ATP1B1", "ATP1B2", "ATP1B3", "ATP1B4", "FXD2"]	0	0	20
PAPP			0	0	0.2
PD-407824				0	20
PD153035 hydrochloride			0.0068	0.0057	19.998253
PF-429242 dihydrochloride			0	0	20
PMEG hydrate				0	20
Palonosetron hydrochloride	["Serotonin receptor antagonist"]	["HTR3A"]	0	0.004	0.02
Paroxetine hydrochloride hemihydrate (MW = 374.83)	["Selective serotonin reuptake inhibitor (SSRI)"]	["SLC6A4", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "HTR2A", "SLC6A2"]	0	0	20.000106
Parthenolide					20
Pazopanib	["KIT inhibitor", "PDGFR receptor inhibitor", "VEGFR inhibitor"]	["KDR", "KIT", "FLT1", "FLT4", "PDGFRB", "PDGFRA", "BRAF", "CSF1R", "CYP2B6", "CYP2C8", "CYP2E1", "DDR2", "FGF1", "FGFR1", "FGFR3", "ITK", "SH2B3"]	0.0037	0	20
Pentamidine		["TRDMT1"]	0	0	20
Pentamidine isethionate		["TRDMT1"]	0		20
Pergolide methanesulfonate	["Dopamine receptor agonist"]	["DRD1", "DRD2", "ADRA2A", "ADRA2B", "ADRA2C", "DRD3", "DRD4", "DRD5", "HTR1A", "HTR1B", "HTR1D", "HTR2A", "HTR2B", "HTR2C", "ADRA1A", "ADRA1B", "ADRA1D", "KCN5"]	0	0.0512	2
Perphenazine	["Dopamine receptor antagonist"]	["DRD2", "CALM1", "DRD1", "HRH1", "HTR2A", "HTR2C", "HTR6", "HTR7"]	0.0039	0.0125	0.02
Phenamil methanesulfonate	["TRPV antagonist"]	["PKD2L1"]	0.0325	0.1496	2
Pheniramine maleate	["Histamine receptor antagonist"]	["HRH1"]	0.0169	0.0024	20
Phorbol 12-myristate 13-acetate	["PKC activator"]	["CD4", "KCNT2", "PRKCA", "TRPV4"]	0	0	2
Pifithrin-mu	["HSP inhibitor"]	["HSPA1A", "TP53"]	0	0.0003	2
Pimozide	["Dopamine receptor antagonist"]	["DRD2", "DRD3", "CACNA1I", "CALM1", "HRH1", "HTR1A", "HTR2A", "KCNA10", "KCNH2"]	0.001	0.0174	20
Piperlongumine	["Glutathione transferase inhibitor"]	String[]	0.0747	0.0144	2
Podophyllotoxin	["Microtubule inhibitor", "Tubulin inhibitor"]	["IGF1R", "CASP3", "TOP2A", "TUBA4A", "TUBB"]	0.001	0	2
Prazosin hydrochloride	["Adrenergic receptor antagonist"]	["ADRA1A", "ADRA1B", "ADRA1D", "ADRA2A", "ADRA2B", "ADRA2C", "CDK1", "KCNH2", "KCNH6", "KCNH7"]	0	0.0003	0.2
Progesterone		["PGR", "CYP17A1", "NR3C2", "CATSPER1", "CATSPER2", "CATSPER3", "CATSPER4", "CYP2C19", "ESR1", "OPRK1", "TRPC5"]	0	0	20

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
Proguanil	["Dihydrofolate reductase inhibitor"]	["CYP2C19", "DHFR"]	0.0016	0	20
Promazine hydrochloride	["Dopamine receptor antagonist"]	["CHRM5", "DRD2", "ADRA1A", "ADRA1B", "ADRA1D", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "DRD1", "DRD3", "DRD4", "HRH1", "HTR2A", "HTR2C"]	0.0243	0	2
Propafenone hydrochloride	["Antiarrhythmic"]	["KCNH2", "SCN5A", "ADRB1", "ADRB2", "KCNA5", "KCNK2", "KCNK3"]	0	0	20
Propionylpromazine hydrochloride			0	0	20
Protriptyline hydrochloride	["Tricyclic antidepressant"]	["SLC6A2", "SLC6A4"]	0	0	20
Psoralidin			0	0	20
Pyridostatin trifluoroacetate salt			0	0.1911	20
Quinacrine dihydrochloride					20
Quinidine sulfate				0	20
RN-9893			0.2081	0.1042	0.02
RU-SKI 43 maleate			0	0	20
Rabeprazole sodium	["ATPase inhibitor", "Gastrin inhibitor"]	["ATP4A", "CYP2C19"]	0.0919	0	20.000126
Raloxifene hydrochloride	["Estrogen receptor antagonist", "Selective estrogen receptor modulator (SERM)"]	["ESR1", "ESR2", "ACVRL1", "ENG"]	0	0	20
Ranolazine dihydrochloride	["Sodium channel blocker"]	["SCN9A", "SCN10A", "SCN5A", "SLC22A2"]	0.0022	0.0209	2
Reserpine	["Vesicular monoamine transporter inhibitor"]	["SLC18A2", "SLC18A1"]	0	0	20
Ro 11-1464			0.0028	0	20
Ro 90-7501	["Beta amyloid inhibitor"]	["APP"]	0	0.0012	20
Roscovitine	["CDK inhibitor"]	["CDK2", "CDK9", "CDK7", "CDK1", "CDK5"]	0	0	20
Rotenone		["MT-ND1"]	0.0004	0	20
Ruthenium red			0.0627	0	20
S-(+)-Fluoxetine hydrochloride			0	0.2889	2
S-Methylisothiurea hemisulfate			0.0518	0.2182	20
SB 202190	["p38 MAPK inhibitor"]	["MAPK14", "AKT1", "ALOX5", "CHEK1", "GSK3B", "LCK", "MAPK1", "MAPK11", "MAPK12", "MAPK8", "PRKCA", "ROCK1", "RPS6KB1", "SGK1"]	0	0	2.0000599
SB 415286	["Glycogen synthase kinase inhibitor"]	["GSK3B", "GSK3A", "RPS6KB1"]	0	0.0012	20
SB743921 hydrochloride			0	0	20
SID 3712249			0.0025	0	20
SKF 83959 hydrobromide			0.0457	0	20
SMER28			0	0	20
SP600125			0	0	20
SR 59230A oxalate	["Adrenergic receptor antagonist"]	["ADRB3", "ADRB1", "ADRB2"]	0.0107	0	20
SR9243			0.0694	0.0012	20
SU 5416			0.0212	0.3581	2

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
SU1498			0	0	20
Sanguinarine chloride					20
Sertaconazole nitrate	["Sterol demethylase inhibitor"]	String[]	0.0119	0	20
Stattic					20
Sunitinib	["FLT3 inhibitor", "KIT inhibitor", "PDGFR receptor inhibitor", "RET tyrosine kinase inhibitor", "VEGFR inhibitor"]	["FLT3", "KDR", "KIT", "FLT4", "FLT1", "PDGFRA", "PDGFRB", "RET", "CSF1R", "FGFR1"]	0	0.3832	20
Supercinnamaldehyde			0	0.0144	20
Suprafenacine			0	0	20
T0070907			0	0	20
TBBz			0.001	0.0026	2
TIC10 angular			0	0	20
Tacrine	["Acetylcholinesterase inhibitor", "Acetylcholine release stimulant", "Butyrylcholinesterase inhibitor", "Potassium channel antagonist"]	["ACHE", "BCHE"]	0	0	20
Tamoxifen citrate					20
Taurine			0	0.4691	20
Testosterone	["Androgen receptor agonist"]	["AR", "CYP19A1", "CYP2C19", "CYP2C8", "CYP3A5"]	0.0126	0.0554	20
Tetraethylthiuram disulfide					20
Thiabendazole			0.0103	0.0012	20
Thiocolchicine			0	0	20
Tirapazamine			0.0007	0.0021	2.2628831
Tizanidine hydrochloride	["Adrenergic receptor agonist"]	["ADRA2A", "ADRA2B", "ADRA2C", "CYP1A2", "NISCH"]	0	0	20
Tolazoline	["Adrenergic receptor antagonist"]	["ADRA2A", "ADRA2B", "ADRA2C", "ADRA1A", "HRH1", "HRH2"]	0.0338	0.1042	20
Topotecan hydrochloride hydrate	["Topoisomerase inhibitor"]	["TOP1", "TOP1MT"]	0	0	2
Torin2			0	0	20
Trifluoperidol hydrochloride			0.0007	0	20
Triflupromazine hydrochloride	["Dopamine receptor antagonist"]	["HTR2B", "CHRM1", "CHRM2", "CHRNA7", "DRD1", "DRD2"]	0	0	20
Trihexyphenidyl		["CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5"]	0.0467	0.1044	0.2
Trilostane			0	0	20
Trimipramine maleate	["Norepinephrine reuptake inhibitor", "Tricyclic antidepressant"]	["SLC6A2", "SLC6A4", "SLC6A3", "ADRA1A", "ADRA1B", "ADRA2A", "ADRA2B", "ADRB1", "ADRB2", "ADRB3", "CHRM1", "CHRM2", "CHRM3", "CHRM4", "CHRM5", "DRD1", "DRD2", "DRD5", "HRH1", "HTR1A", "HTR1D", "HTR2A", "HTR2C", "HTR3A"]	0.0216	0	20
Tyrphostin AG 879			0	0	20
U-101958 maleate			0.3587	0.1733	2

Compound name	MOA	Targets	RMPV750	RMPV1500	Conc.
U-62066			0.5627	0.0015	20
U0126		["JAK2", "MAP2K1", "MAP2K2", "MAP3K1", "MAP3K2"]	0	0	20
UNC0379 trifluoroacetate salt			0.0013	0	2
Vinblastine	["Microtubule inhibitor", "Tubulin inhibitor"]	["TUBB", "JUN", "TUBA1A", "TUBD1", "TUBE1", "TUBG1"]	0	0.1859	20
Vincristine sulfate		["TUBB", "TUBA4A"]	0.0004	0	20
Vorinostat					20
WIN 62,577			0.0101	0.0021	2
WZ4003			0	0	0.2
Wiskostatin	["Neural Wiskott-Aldrich syndrome protein inhibitor"]	["WAS", "WASL"]	0	0	20
XL388			0	0	20
Y-27632 dihydrochloride	["Rho associated kinase inhibitor"]	["ROCK1", "ROCK2", "LRRK2", "PKIA", "PKN2", "PRKACA", "PRKCE"]	0.0268	0.0057	0.2
Yoda1			0	0.0443	2
alpha-Lobeline hydrochloride			0.0107	0.0347	0.02
beta-Lapachone					20

Supp. Table 1: Compound list in chemical HCS experiment. Description of the name of the screened compounds (CompoundName), their mechanisms of action (MOA) and genetic targets (Targets), their robust morphological perturbation value (FDR-corrected p-value for a comparison to matching DMSO controls) computed on the plate seeded with 750 cells (RMPV750) or 1500 cells (RMPV1500), as well as their concentration in μM (Conc.). For a machine-readable version of this table, see FigShare dataset (DOI: 10.1101/2021.06.18.448961). For accessing the data programatically see the GitHub repository (DOI: 10.5281/zenodo.5659932).

Rank	Feature	Type	Object
1	Granularity_10_CorrCM_median	Granularity, CellMask	Cell
2	Granularity_9_CorrCM_median	Granularity, CellMask	Cell
3	AreaShape_FormFactor_1_median	Shape	Nucleus
4	AreaShape_Solidity_median	Shape	Cell
5	AreaShape_Zernike_3_1_1_median	Shape	Cell
6	Intensity_MassDisplacement_CorrDNA_1_median	Intensity distribution, DAPI	Nucleus
7	Granularity_7_CorrActin_median	Granularity, Phalloidin	Cell
8	AreaShape_Compactness_median	Shape	Cell
9	AreaShape_Compactness_1_median	Shape	Nucleus
10	RadialDistribution_RadialCV_CorrActin_2of3_median	Intensity distribution, Phalloidin	Cell

Supp. Table 2: Top 10 most variable features for the plate seeded with 750 cells per well. Features are ranked by decreasing median absolute deviation. As the data is normalized on the DMSO controls, this corresponds to the features varying the most in response to chemical perturbations across the plate.

MOA	Feature 1	Feature 2	Feature 3	Feature 4
Glycogen synthase kinase inhibitor	RadialDistribution_ZernikeMagnitude_CorrDNA_7_3_median	AreaShape_Zernike_8_6_1_median	RadialDistribution_ZernikeMagnitude_CorrCM_9_3_median	Intensity_MADIntensity_CorrCM_median
CDK inhibitor	AreaShape_Zernike_8_6_1_median	RadialDistribution_ZernikeMagnitude_CorrCM_9_3_median	Intensity_MaxIntensity_CorrCM_median	AreaShape_Zernike_4_0_1_median
Histamine receptor antagonist	AreaShape_Zernike_4_2_2_median	AreaShape_Zernike_4_0_1_median	RadialDistribution_ZernikeMagnitude_CorrCM_9_3_median	Intensity_MaxIntensity_CorrCM_median
Selective serotonin reuptake inhibitor	AreaShape_Zernike_4_2_2_median	AreaShape_Zernike_4_0_1_median	RadialDistribution_ZernikeMagnitude_CorrDNA_7_3_median	Granularity_1_CorrActin_median
Norepinephrine reuptake inhibitor	AreaShape_Zernike_4_2_2_median	AreaShape_Zernike_4_0_1_median	RadialDistribution_ZernikeMagnitude_CorrDNA_7_3_median	AreaShape_Zernike_9_3_2_median

Supp. Table 3: Top 4 most distinctive features (absolute log-fold change of mean field-of-view profiles) as compared to DMSO controls, for selected Mechanisms Of Actions (MOAs) for the plate seeded with 750 cells per well. Recurrent features are colored consistently across MOAs.

Supplementary material - Morphological profiling of human T and NK lymphocytes by high-content cell imaging

This section reproduces the supplementary material corresponding to the paper entitled “Morphological profiling of human T and NK lymphocytes by high-content cell imaging” presentend in section 3.2.

Cell Reports, Volume 36

Supplemental information

Morphological profiling of human T and NK

lymphocytes by high-content cell imaging

Yolla German, Loan Vulliard, Anton Kamnev, Laurène Pfajfer, Jakob Huemer, Anna-Katharina Mautner, Aude Rubio, Artem Kalinichenko, Kaan Boztug, Audrey Ferrand, Jörg Menche, and Loïc Dupré

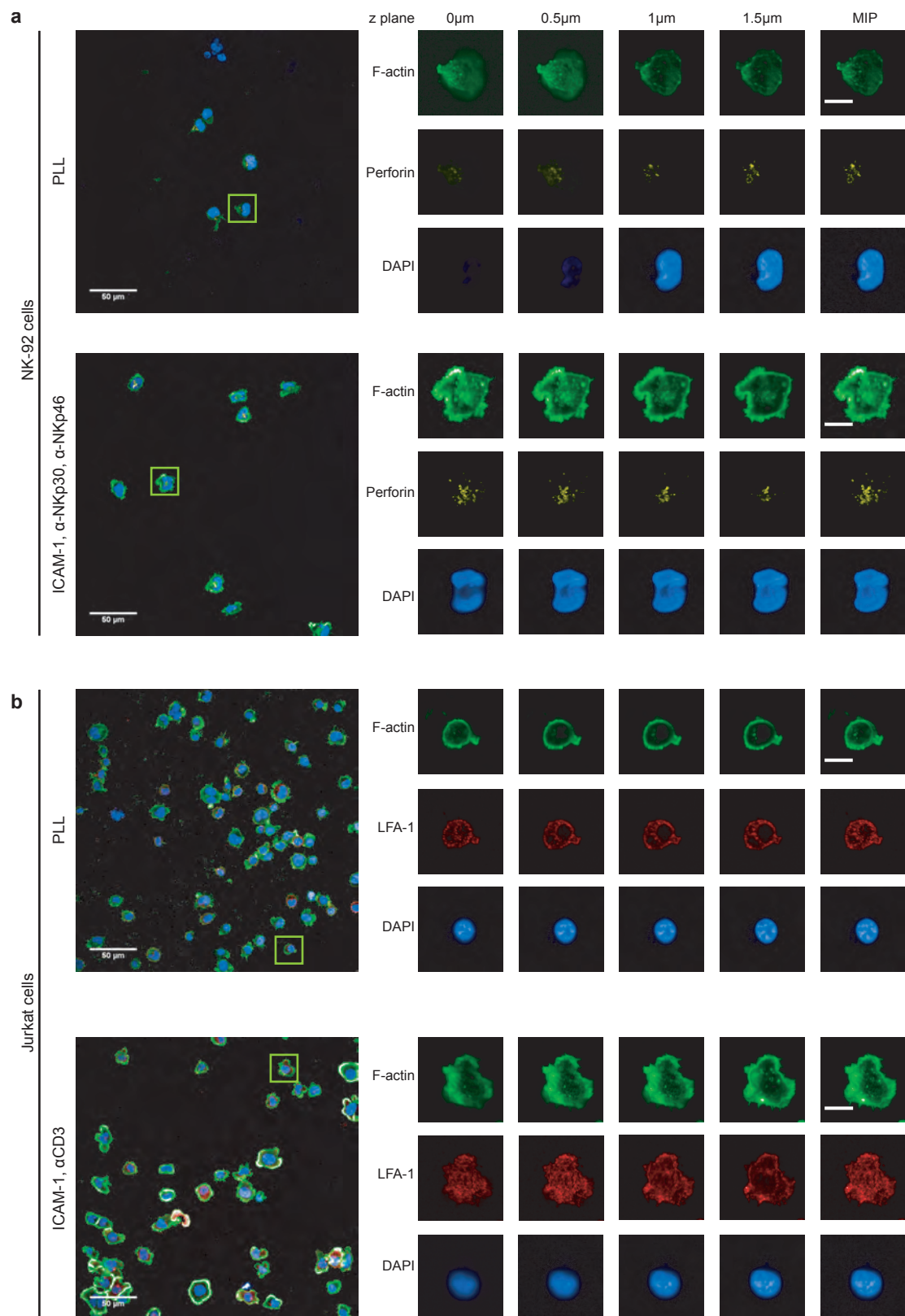


Figure S1 | Individual channels and z planes of single cells isolated from representative fields of view. a. Maximum intensity projection (MIP) of a representative field of view of NK-92 cells seeded on PLL (top) or ICAM-1, anti-NKp30 and NKp-46 (bottom), with zoom on a single representative cell stained for F-actin (green), perforin granules (yellow) and nuclei (DAPI) imaged at 4 z-planes with a step of 0.5 μ m and its MIP. **b.** MIP of a representative field of view of Jurkat cells seeded on PLL (top) ICAM-1, anti-CD3 (bottom), with zoom on a single cell stained for F-actin (green), LFA-1 (red) and nuclei (DAPI) and imaged at 4 z-planes with a step of 0.5 μ m and its MIP. Scale bars: Field of view 50 μ m and single cell 10 μ m. Related to Figure 1.

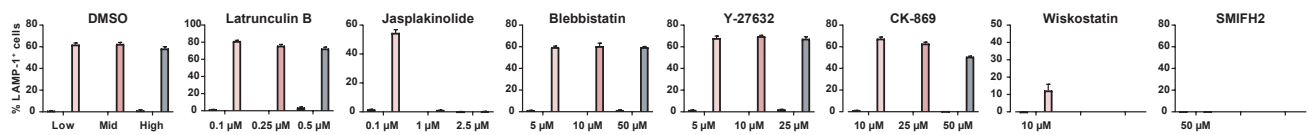


Figure S2 | Actin-targeting drugs differentially affect lytic granule exocytosis in NK-92 cells. Percentage of LAMP-1⁺ NK-92 cells upon stimulation with coated ICAM-1 and anti-NKp30 Ab. DMSO or actin-targeting drugs were tested at the indicated concentrations. Histograms correspond to mean \pm SD of triplicate measurements. Related to Figure 2.

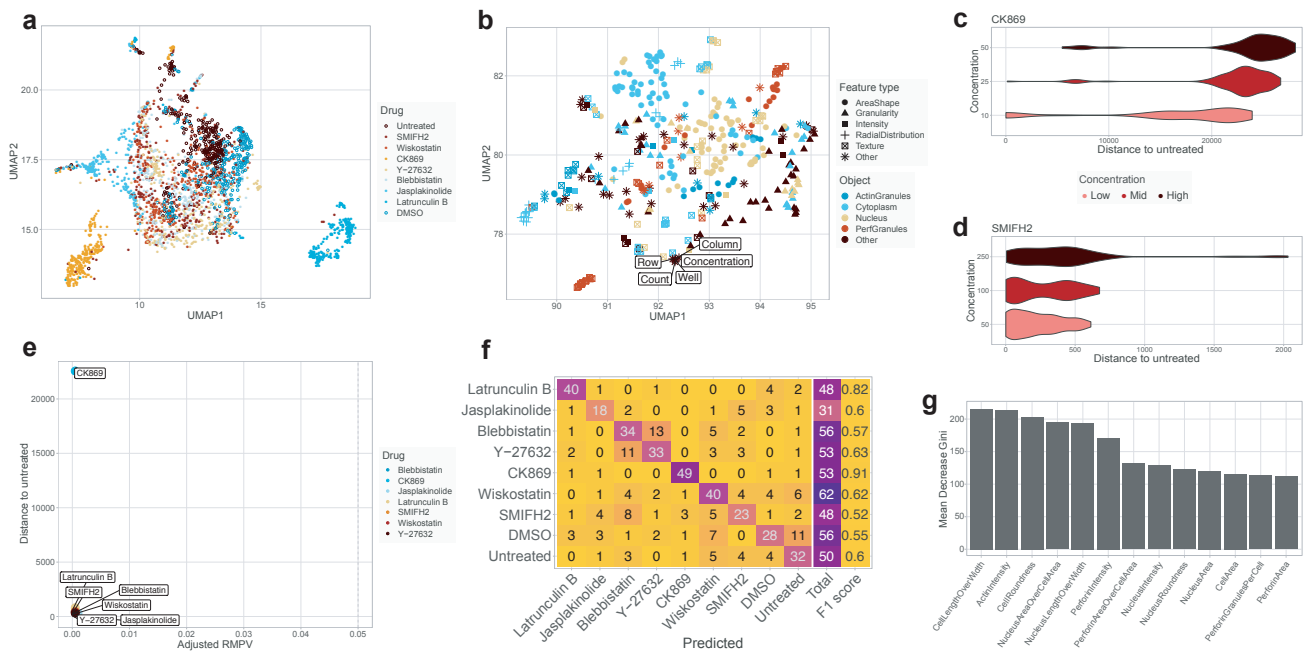


Figure S3 | Drug treatment leads to distinct immunological synapse phenotypes in NK-92 cells. **a.** Confusion matrix and class-wise performance of a random forest model trained to predict drug treatment based on 13 curated morphological descriptors of NK-92 seeded on ICAM-1, anti-NKp30 and anti-NKp46. **b.** Importance of the 13 morphological descriptors for the classification described in panel (a). **c.** UMAP representing the clustering of all the drugs and the untreated conditions. **d.** UMAP representing the relations between confounders and morphological descriptors, obtained by fitting the UMAP on the transpose of the data underlying panel (c). **e-f.** Violin plots representing the effect size of drug concentrations on morphological descriptors for (e) CK-869 and (f) SMIFH2. **g.** FDR-corrected Robust Morphological Perturbation Value (RMPV) of the different drugs. Related to Figure 3.

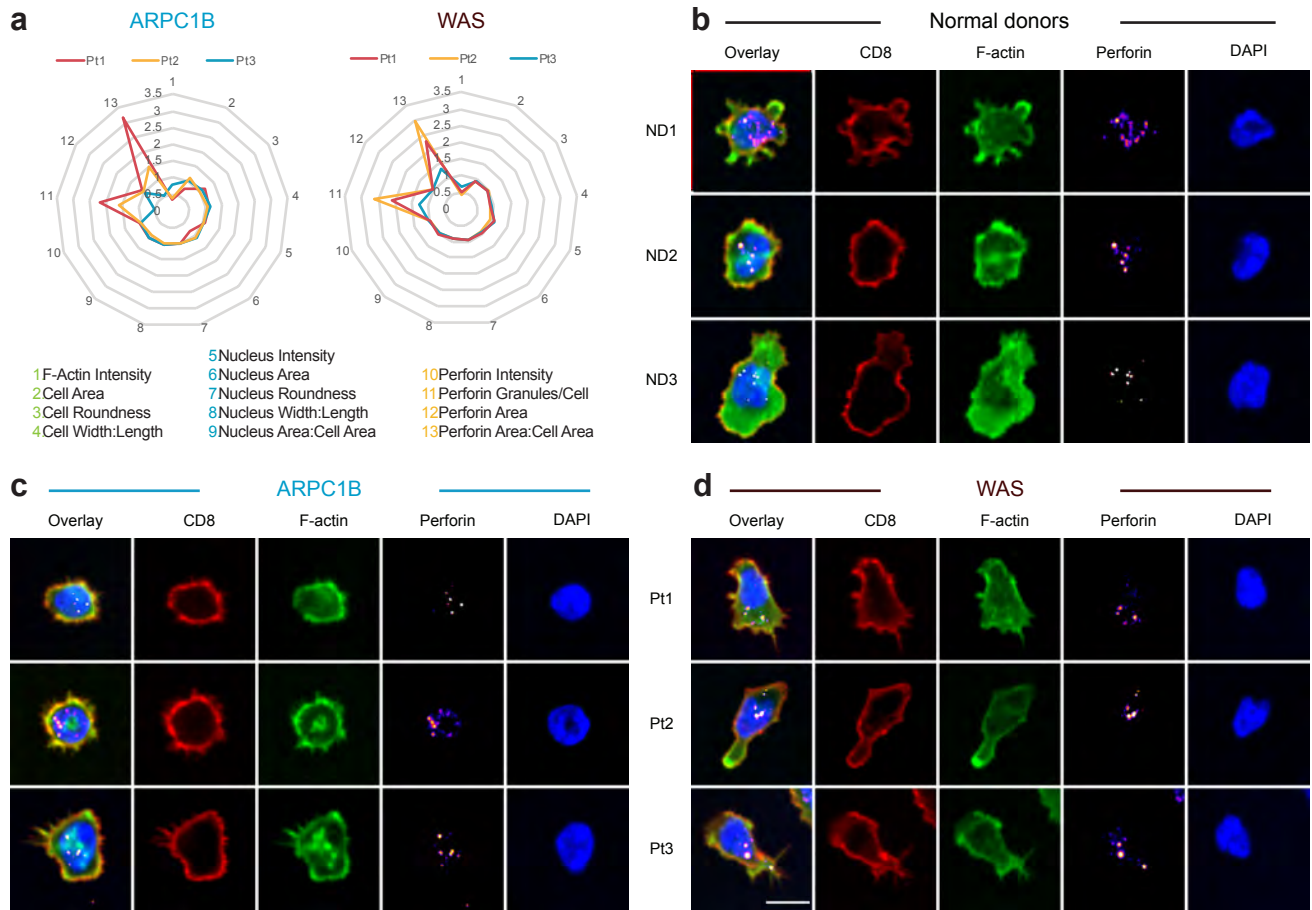


Figure S4 | Individual morphology of patient CD8⁺ T cells. **a.** Average characteristics of the IS of CD8⁺ T cells from patients represented as fold change with respect to the average of three normal donors (312-3091 cells per donor or patient). **b-d.** Representative images of CD8⁺ T cells from **(b)** normal donors, **(c)** ARPC1B-deficient or **(d)** WASP-deficient patients. Scale bar: 10 μ m. Related to Figure 5.

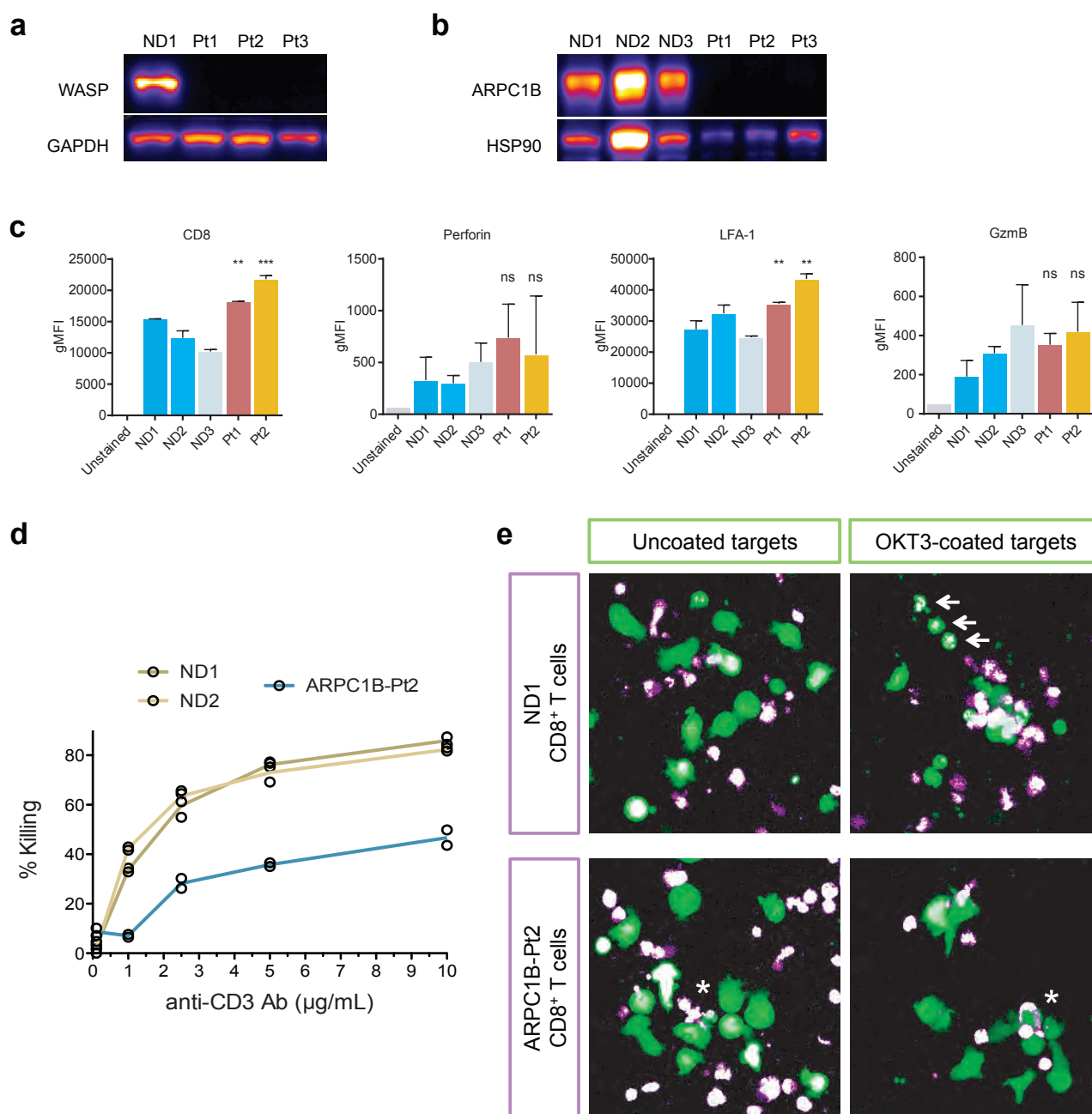


Figure S5 | Phenotypic and functional characterization of CD8⁺ T cells **a-b.** Western blot analysis of WASP AND ARPC1B expression in the corresponding patients. **c.** Expression levels of CD8, perforin, LFA-1 and granzyme B in the indicated T cells. **d.** Cytotoxic activity of CD8⁺ T cells against anti-CD3-coated P815 target cells. **e.** Representative images of the interaction of CD8⁺ T cells (purple) with P815 target cells (green). Arrows point to dead target cells. Stars point to ARPC1B-deficient T cells with aberrant morphologies. Related to figure 5 and 6.

	NK-92		
	PLL	ICAM-1, α -NKp30, α -NKp46	Fold increase
F-Actin intensity	0.065098277	0.110223848	1.69319149
Cell Area (μm^2)	2469.321237	3130.618434	1.26780525
Cell Roundness	0.674079624	0.444435641	0.65932217
Cell Width:Length	0.686074044	0.71563093	1.04308119
Nucleus Intensity	0.067203031	0.067329026	1.00187485
Nucleus Area(μm^2)	1037.718811	1333.179255	1.2847211
Nucleus Roundness	0.706061099	0.689874902	0.97707536
Nucleus Width:Length	0.685545171	0.706542199	1.03062822
Nucleus Area:Cell Area	0.427455169	0.431859895	1.01030453
Perforin Intensity	0.018403047	0.018502314	1.00539403
Perforin Granules/Cell	13.04941437	16.7196273	1.28125499
Perforin Area(μm^2)	86.05831253	117.5449158	1.36587521
Perforin Area:Cell Area	0.035039484	0.036882184	1.05258922

Table S1 | Mean values and fold increase of immunological synapse parameters in NK-92 cells. Mean values of individual parameters pertaining to the immunological synapse in NK-92 cells seeded on PLL or ICAM-1, anti-NKp30 and anti-NKp46, and the fold change of the ratio of each mean value on the stimulated condition with respect to PLL. Intensity is measured in arbitrary units and area in μm^2 . Related to Figure 1.

	Jurkat		
	PLL	ICAM-1, α -CD3	Fold increase
F-Actin intensity	0.000289437	0.000437181	1.5104544
Cell Area(μm^2)	1921.009837	2690.327023	1.4004754
Cell Roundness	0.530654443	0.61184158	1.15299436
Cell Width:Length	0.800192404	0.767932394	0.95968468
Nucleus Intensity	0.07313481	0.074429279	1.01769975
Nucleus Area(μm^2)	879.0221429	995.945146	1.13301486
Nucleus Roundness	0.830340643	0.806271514	0.97101295
Nucleus Width:Length	0.779080494	0.754652368	0.96864493
Nucleus Area:Cell Area	0.460560432	0.379756314	0.82455263
LFA-1 Intensity	0.006679646	0.007880989	1.17985131
LFA-1 Area(μm^2)	1728.306565	2526.153065	1.46163483
LFA-1 Area:Cell Area	0.900646361	0.938478533	1.04200558

Table S2 | Mean values and fold increase of immunological synapse parameters in Jurkat cells. Mean values of individual parameters pertaining to the immunological synapse in Jurkat cells seeded on PLL or ICAM-1 and anti-CD3, and the fold change of the ratio of each mean value on the stimulated condition with respect to PLL. Intensity is measured in arbitrary units and area in μm^2 . Related to Figure 1.

Normal donors	Estimate	Std. Error	t value	Pr(> t)
Intercept	-0.205921	0.1595205	-1.291	0.196754
AreaShape_Perimeter	-0.004858	0.0005022	-9.674	< 2e-16
AreaShape_MaximumRadius	0.4582657	0.0158259	28.957	< 2e-16
AreaShape_MinorAxisLength	-0.0095155	0.0030981	-3.071	0.002132
AreaShape_MeanRadius	0.3306122	0.0440017	7.514	5.86E-14
AreaShape_FormFactor	-0.7657327	0.2011575	-3.807	0.000141
Intensity_MeanIntensity_CorrActin	7.0413384	0.585683	12.022	< 2e-16
RadialDistribution_FracAtD_CorrActin1_1of3	-3.0109743	0.4111895	-7.323	2.47E-13
RadialDistribution_FracAtD_CorrActin1_2of3	3.8917133	0.2652633	14.671	< 2e-16
Adjusted R-squared: 0.2884				

ARPC1B-Pt1	Estimate	Std. Error	t value	Pr(> t)
Intercept	0.8029111	0.2525619	3.179	0.00148
AreaShape_Perimeter	-0.0049727	0.0009152	-5.433	5.61E-08
AreaShape_MaximumRadius	0.5992636	0.0273474	21.913	< 2e-16
AreaShape_MinorAxisLength	-0.0060863	0.0072368	-0.841	0.400351
AreaShape_MeanRadius	0.0203988	0.0722717	0.282	0.777754
AreaShape_FormFactor	-0.751234	0.2694728	-2.788	0.005313
Intensity_MeanIntensity_CorrActin	5.0379313	1.3418843	3.754	0.000174
RadialDistribution_FracAtD_CorrActin1_1of3	-0.7157633	0.4704314	-1.522	0.128152
RadialDistribution_FracAtD_CorrActin1_2of3	1.3439737	0.3781638	3.554	0.000381
Adjusted R-squared: 0.2878				

ARPC1B-Pt2	Estimate	Std. Error	t value	Pr(> t)
Intercept	0.0541722	0.180384	0.3	0.763939
AreaShape_Perimeter	-0.004052	0.0004741	-8.546	< 2e-16
AreaShape_MaximumRadius	0.5642439	0.0207764	27.158	< 2e-16
AreaShape_MinorAxisLength	-0.0157821	0.0043454	-3.632	0.000282
AreaShape_MeanRadius	0.11303	0.0553829	2.041	0.041274
AreaShape_FormFactor	-1.1368577	0.2379435	-4.778	1.78E-06
Intensity_MeanIntensity_CorrActin	6.7594643	1.1084589	6.098	1.09E-09
RadialDistribution_FracAtD_CorrActin1_1of3	2.5718437	0.5322001	4.832	1.36E-06
RadialDistribution_FracAtD_CorrActin1_2of3	1.4839242	0.352878	4.205	2.62E-05
Adjusted R-squared: 0.324				

Table S3 | Coefficient values in linear models of the distance between lytic granules and cell edges based on morphological descriptors in single lymphocytes. Related to Figure 6.

Donor	Row	Column	Field	X	Y	Mean minimum distance granule-edge at z=1	Cell perimeter	Cell maximum radius	Cell minor axis length	Cell mean radius	Cell roundness	Mean intensity F-actin	Radial distribution at z=1, bin 1/3	Radial distribution at z=1, bin 2/3
ND3	14	8	9	387.75	230.75	4.61511	193.924	18.1108	39.2498	6.61522	0.56706	0.068513	0.11376	0.300318
ND1	3	8	17	734.60385	602.26538	9.81357	198.888	19.0263	42.6223	6.79265	0.567696	0.0646902	0.0723923	0.277189
ND2	4	15	32	172	406.5	13.3571	212.409	17.4642	42.2882	6.60718	0.546465	0.0654103	0.0497074	0.271457
ARPC1B-Pt2	10	13	22	334.70238	591.58333	12.0592	228.859	18.6011	41.7853	6.29407	0.394436	0.0539251	0.146708	0.29765
ARPC1B-Pt2	9	14	27	394.80949	975.90586	8.93749	254.38	18.6815	41.0687	6.2795	0.377715	0.0587798	0.0765422	0.280822
ARPC1B-Pt2	9	14	18	671.72049	157.47093	7.95421	232.001	19	40.6662	6.31474	0.391061	0.0525494	0.0407361	0.276465

Table S4 | Morphological descriptors values of representative cells. The cells are described in the same order as shown in Figure 6b. In the CellProfiler analysis, the corresponding parameters in the table are respectively called 'Mean_FilterNKPerfGranules1_Location_Center_X', 'Mean_FilterNKPerfGranules1_Location_Center_Y', 'Mean_FilterNKPerfGranules1_Distance_Minimum_FilterNKCytoplasm', 'AreaShape_Perimeter', 'AreaShape_MaximumRadius', 'AreaShape_MinorAxisLength', 'AreaShape_MeanRadius', 'AreaShape_FormFactor', 'Intensity_MeanIntensity_CorrActin', 'RadialDistribution_FracAtD_CorrActin1_1of3' and 'RadialDistribution_FracAtD_CorrActin1_2of3'. Related to Figure 6.

Curriculum vitae

This final section includes my curriculum vitae, detailing my education and work experiences, as well as an up-to-date list of publications, all arising from my doctoral work.

Loan Vulliard

Curriculum Vitae

Julius-Tandler-Platz 2/15
1090, Vienna, Austria
☎ (+43)660 2877564
✉ lvulliard@cemm.at
<http://vulliard.loan/>
French citizenship

Computational approaches to human health and diseases

Education

- 2017–Present **PhD candidate - Medical Informatics, Biostatistics & Complex Systems, CeMM**
Research Center for Molecular Medicine & Max Perutz Labs, Vienna, Austria.
Training and research in the field of molecular biology for human health, using computational approaches to tackle post-genomic challenges.
- 2012–2017 **Master's Degree in Bioscience - Bioinformatics & Modeling, INSA Lyon, Villeurbanne, France.** Engineering university.
Biology, computer science, mathematics and transdisciplinary courses. Two years in the English-speaking section, then in the Music studies section.
- Fall 2016 **Department of Biological Sciences, KAIST, Daejeon, Republic of Korea.**
Academic exchange, as part of my studies at INSA Lyon.
- 2009–2012 **BAC S (scientific A-level), Lycée Gustave Eiffel, Bordeaux, France.**
High school diploma with distinctions (17.76/20), European section.

Experience

- Since September 2020 **Network medicine and perturbation biology, Max Perutz Labs, Vienna BioCenter, Vienna, AUSTRIA.**
Predoctoral fellow in the **Jörg Menche group**, developing computational approaches to understand and interpret the large datasets derived from post-genomic technologies.
- Arrayed high-content screening of interactions between drugs and gene knockouts
 - Network framework to infer rules underlying the cumulative effect of internal and external perturbations in biological systems
- October 2017 - August 2020 **Network medicine and perturbation biology, CeMM, Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, AUSTRIA.**
Joined the **research group of Prof. J Menche** before its move to the Max Perutz Labs.
- January - June 2017 **Applied bioinformatics in cancer research, DTU Bioinformatics, Technical University of Denmark, Copenhagen, DENMARK.**
Research internship in the **Machine learning and immunoinformatics** team, led and supervised by Prof. M Nielsen, co-supervised by Dr. A C Eklund.
- Studied neoantigen profiles in cancer patients across the TCGA database
 - Studied machine learning algorithms for bioinformatics
- May - August 2016 **Applied bioinformatics in cancer research, Institut de Recherche en Cancérologie de Montpellier (IRCM), INSERM, Montpellier, FRANCE.**
Research internship in the **Cancer bioinformatics and systems biology** team led by Prof. J Colinge, under the supervision of Prof. E Cornillot.
- Studied the association rate between HLA alleles and neo-antigens in tumours
 - Designed and developed an HLA typing pipeline

- 2014–2015 **Communication Manager**, RAID INSA LYON - ORANGE, BdE INSA Lyon, Villeurbanne, FRANCE.
 Outdoor sport event : 200 runners, 3 days, 25 student organizers working on it yearly with a budget of approximately 50k€. Part of the biggest Students' Union in France.
 ○ **Managed a team of 4 people**
 ○ Realized graphic productions, articles, videos, and other communication material
- June - August 2013 **Line Worker**, SBE LTD., Ashford, ENGLAND.
 ○ Experienced industry and line work in a factory, ran diagnoses on cell phones

Skills

Programming

- Advanced **Python, R, Julia**
 Intermediate Bash, C, SQL
 Basic HTML, PHP, Java, JavaScript, C++, C#, Matlab

Mathematics

- Intermediate **Data Science, Machine Learning, Networks Science**, Modeling, Biostatistics, Differential Calculus, Linear Algebra, Probability
 Basic Analysis, Graph Theory, Game Theory

Biology

- Intermediate **Screening, Imaging**, Genetics, Immunology, NGS, Genomics, Transcriptomics
 Basic Cell Biology, Cancer Biology, Epigenetics, Proteomics, Pharmacology

Languages

- Advanced French (native), English (C1 European level, TOIEC certification in 2016)
 Intermediate German (C1 European level, ÖSD certification in 2021)
 Basic Russian, Korean

Other

- Intermediate Unix, Use of computer clusters, GitHub Actions, L^AT_EX, RegEx, Adobe Suite

Achievements

- 5 peer-reviewed publications in international journals and multiple preprints - see Publication list
- Julia package - BioProfiling.jl - Available on the Julia Registries
- R package - BioCircos - Available on CRAN - More than 20000 downloads
- Poster prize - Young Scientists Retreat of Institut Curie and CeMM - Network-based Approach to Drug-gene Interactions - Bratislava, 09/2018
- Poster prize - EMBO/FEBS/Institut Curie Course on Computational Systems Biology of Cancer - Understanding Chemical-Genetic Interactions - 09/2020
- James Taylor Foundation's JTech scholarship - 11/2020
- Travel grant - awarded to present at the JOBIM conference in Nantes, 07/2019
- Competitive programming - Bioinformatics Contest 2017 and 2019: Top 5 %
- Competitive programming - Google Hash Code 2017: 2nd best team in Denmark

List of publications

Journal articles

- **Vulliard, L.**, Hancock, J., Kamnev, A., Fell, C.W., Ferreira da Silva, J., Loizou, J.I., Nagy, V., Dupré, L., Menche, J., 2021. BioProfiling.jl: profiling biological perturbations with high-content imaging in single cells and heterogeneous populations. **Bioinformatics**. <https://doi.org/10.1093/bioinformatics/btab853>
- German, Y.*, **Vulliard, L.***, Kamnev, A., Pfajfer, L., Huemer, J., Mautner, A.-K., Rubio, A., Kalinichenko, A., Boztug, K., Ferrand, A., Menche, J., Dupré, L., 2021. Morphological profiling of human T and NK lymphocytes by high-content cell imaging. **Cell Reports** 36, 109318. <https://doi.org/10.1016/j.celrep.2021.109318>
* Equal contribution.
- Brunner J.S., **Vulliard L.**, Hofmann M., Kieler M., Lercher A., Vogel A., Russier M., Brüggenthies J., Kerndl M., Saferding V., Niederreiter B., Junza A., Frauenstein A., Scholtysek C., Mikami Y., Klavins K., Krönke G., Bergthaler A., O'Shea J.J., Weichhart T., Meissner F., S. J., Cheng P., Yanes O., Menche J., J. P., Sharif O., Blüml S., Schabbauer G., 2020. Environmental Arginine Controls Multinuclear Giant Cell Metabolism and Formation. **Nature Communications** 11, 431. <https://doi.org/10.1038/s41467-020-14285-1>
- Schick, S., Rendeiro, A.F., Runggatscher, K., Ringler, A., Boidol, B., Hinkel, M., Májek, P., **Vulliard, L.**, Penz, T., Parapatics, K., Schmidl, C., Menche, J., Boehmelt, G., Petronczki, M., Müller, A.C., Bock, C., Kubicek, S., 2019. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalties in human cancers. **Nature Genetics** 51, 1399–1410. <https://doi.org/10.1038/s41588-019-0477-9>
- Schischlik, F., Jäger, R., Rosebrock, F., Hug, E., Schuster, M., Holly, R., Fuchs, E., Feenstra, J.D.M., Bogner, E., Gisslinger, B., Schalling, M., Rumi, E., Pietra, D., Fischer, G., Faé, I., **Vulliard, L.**, Menche, J., Haferlach, T., Meggendorfer, M., Stengel, A., Bock, C., Cazzola, M., Gisslinger, H., Kralovics, R., 2019. Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. **Blood** 134, 199–210. <https://doi.org/10.1182/blood.2019000519>

Book chapter

- **Vulliard, L.**, Menche, J., 2021. Complex Networks in Health and Disease, in: **Systems Medicine**. Elsevier, pp. 26–33. <https://doi.org/10.1016/B978-0-12-801238-3.11640-X>

Conference abstract

- Schischlik, F., Jäger, R., Rosebrock, F., Hug, E., Schuster, M., Holly, R., Fuchs, E., Milosevic Feenstra, J.D., Bogner, E., Gisslinger, B., Schalling, M., Rumi, E., Pietra, D., Fischer, G., Faé, I., **Vulliard, L.**, Menche, J., Haferlach, T., Meggendorfer, M., Stengel, A., Bock, C., Cazzola, M., Gisslinger, H., Kralovics, R., 2018. Mutational Landscape of the Transcriptome Offers a Rich Neoantigen Resource for Immunotherapy of Myeloproliferative Neoplasms. **Blood** 132, 3058–3058. <https://doi.org/10.1182/blood-2018-99-118049>

Unpublished preprints

- Wang, X.-W., Madeddu, L., Spirohn, K., Martini, L., Fazzone, A., Becchetti, L., Wytock, T.P., Kovács, I.A., Balogh, O.M., Benczik, B., Pétervári, M., Ágg, B., Ferdinandy, P., **Vulliard, L.**, Menche, J., Colonnese, S., Petti, M., Scarano, G., Cuomo, F., Hao, T., Laval, F., Willems, L., Twizere, J.-C., Calderwood, M.A., Petrillo, E., Barabási, A.-L., Silverman, E.K., Loscalzo, J., Velardi, P., Liu, Y.-Y., 2021. Assessment of community efforts to advance computational prediction of protein-protein interactions. **bioRxiv**. <https://doi.org/10.1101/2021.09.22.461292>

List of conference presentations

- Images to Knowledge 2022
- VBC PhD Symposium 2021
- Perutz Symposium 2021
- Crick BioImage Analysis Symposium 2021
- CIRM Research School in Networks and Molecular Biology 2020
- CRISPR and beyond 2020
- EMBO/FEBS/Institut Curie Lecture Course on Computational Systems Biology of Cancer 2020
- EMBO-EMBL Symposium on Systems Genetics 2019
- CytoData 2019
- Mediterranean School of Complex networks 2018
- German Conference on Bioinformatics 2018
- Single Cell Biology 2022
- CRISPR and beyond 2021
- JuliaCon 2021
- CytoData 2020
- NetSci 2020
- CSHL Biological Data Science Conference 2020
- JOBIM 2019
- Young Scientists Retreat of Institut Curie and CeMM 2018
- CeMM Recess 2018

Scientific outreach and activities

- CytoData Society - Resource officer - 2021-2022
- Pint of Science organiser - Vienna - 2019-2021
- PhD student representative - CeMM - 2017-2019