MEDICAL UNIVERSITY OF VIENNA

A Survey of Solute Carrier-Drug Associations

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

Doctor of Philosophy

Submitted by

Adrián César Razquin, MSc.

Supervisor:

Prof. Dr. Giulio Superti-Furga

Center for Physiology and Pharmacology,
Medical University of Vienna
Währinger Straße 13A/HP, 1090 Vienna, Austria

CeMM Research Center for Molecular Medicine
of the Austrian Academy of Sciences
Lazarettgasse 14, AKH BT 25.3, 1090 Vienna, Austria

Vienna, 02/2019

Declaration

The following doctoral thesis is presented in a cumulative format and includes both published and unpublished results. The work described in this thesis has been performed by the author in the group of Prof. Dr. Giulio Superti-Furga at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, unless otherwise stated. A short-term internship in the group of Prof. Dr. Julio Saez-Rodriguez at the Joint Research Center for Computational Biomedicine (JRC-COMBINE), RWTH Aachen, was carried out as part of a collaboration leading to manuscript #2. The author of this thesis is first author in manuscripts #1 and #2 and co-author in manuscript #3. Individual contributions of all authors are listed in detail in the prologue and interlude sections that precede the manuscripts. All other parts of this thesis were written solely by the author with input and feedback provided by Enrico Girardi, Ulrich Goldmann and Giulio Superti-Furga. The publications included in this thesis are:

Manuscript #1, published as:

> *A Call for Systematic Research on Solute Carriers*
> Adrián César-Razquin, Berend Snijder, Tristan Frappier-Brinton, Ruth Isserlin, Gergely Gyimesi, Xiaoyun Bai, Reinhart A. Reithmeier, David Hepworth, Matthias A. Hediger, Aled M. Edwards, and Giulio Superti-Furga
> Cell. 2015 Jul 30; 162(3):478-87. doi: 10.1016/j.cell.2015.07.022

Manuscript #2, published as:

> *In silico Prioritization of Transporter–Drug Relationships From Drug Sensitivity Screens*
> Adrián César-Razquin, Enrico Girardi, Mi Yang, Marc Brehme, Julio Saez-Rodriguez, and Giulio Superti-Furga
> Frontiers in Pharmacology. 2018 Sep 7; 9:1011. doi: 10.3389/fphar.2018.01011

Manuscript #3, unpublished, provided here as a draft for subsequent submission:

> *A Functional Survey of Solute Carrier-Drug Interactions*
> Enrico Girardi, Adrián César-Razquin, Konstantinos Papakostas, Sabrina Lindinger, Justyna Konecka, Jennifer Hemmerich, Stefanie Kickinger, Giuseppe Fiume, Anna Ringler, Charles-Hugues Lardeau, Richard Kumaran Kandasamy, Stefan Kubicek, Gerhard Ecker, and Giulio Superti-Furga

Table of contents

List of Figures and Tables

Abstract

The transport mode of most drugs and drug-like compounds across cellular membranes is currently not fully resolved. Yet, although the relative contributions of passive and carrier-mediated processes remain a matter of controversy, the role played by membrane transporters is increasingly recognized. Two main protein superfamilies are known to be involved in drug transport: the ATP-Binding Cassette (ABC) family of efflux transporters, relatively well-characterized and often involved in multidrug resistance, and the larger, less-studied and more heterogeneous group of Solute carriers (SLCs). Literature analysis showed that SLCs represent a highly neglected gene group, presenting the largest publication asymmetry of all human protein families, and including an important proportion of poorly characterized and orphan members. Similarly, a compilation of direct drug-transporter relationships also showed that most cases concentrated around a few well-characterized SLC families, with other members being only exceptionally involved. A closer, systematic inspection of the SLC family of transporters from the standpoint of drug disposition might therefore constitute a key approach to unravel the transport mechanism and dependencies of important therapeutic agents.

This thesis provides a systematic interrogation of SLC-drug associations based on a combination of computational and experimental approaches. On the one hand, we built predictive models of drug sensitivity based on SLC and ABC molecular features (gene expression, copy number variations –CNVs–, single nucleotide variants –SNVs–) using the most comprehensive pharmacogenomics dataset to date, which involves 1,000 annotated cancer cell lines and their sensitivity to 265 compounds. On the other hand, we designed a CRISPR-based genetic screening approach and analysis pipeline that enabled us to experimentally survey transporter dependencies for a set of 60 diverse cytotoxic agents. Both approaches led to the successful identification of known as well as previously undescribed transporter-drug relationships, including direct transport cases and indirect effects, some of which we validated.

The results presented here provide an unbiased prioritization means of potentially pharmacologically relevant interactions, and offer new insights into the transport mode and pharmacodynamics of a number of compounds of great therapeutic relevance. Such

understanding is of primary importance in order to increase the specificity and efficiency of drug therapies and move towards precision medicine approaches.

Zusammenfassung

Der genaue Transportweg welchen pharmazeutische Wirkstoffe sowie andere medikamenten-ähnliche niedermolekulare Substanzen benutzen um Zellmembranen zu durchqueren ist bis dato noch unzureichend verstanden. Die Bedeutung welche Membrantransporter in diesem Zusammenhang innehaben wird zunehmend erkannt, auch wenn der relative Anteil von passivem versus Carrier-vermitteltem Transport nach wie vor kontrovers diskutiert wird. Bekanntermaßen sind zwei große Protein-Superfamilien am Wirkstofftransport beteiligt: Die Efflux-Transporter der ATP-Binding Cassette (ATP) Familie, welche verhältnismäßig gut charakterisiert und oft ursächlich an der Entwicklung von Mehrfachresistenzen beteiligt sind, sowie die größere, jedoch weniger genau untersuchte und heterogenere Gruppe der Solute carriers (SLCs). Die genaue Auswertung der existierenden Literatur wies SLCs als eine bislang stark außer Acht gelassene Gengruppe aus. Im Vergleich zu allen anderen humanen Proteinfamilien weisen Mitglieder der SLCs die größte Asymmetrie bezüglich Erwähnung in Publikationen auf. Des Weiteren enthält die SLC-Familie einen großen Teil an noch schlecht charakterisierten Mitgliedern mit unbekannten Transportsubstraten. Ebenso zeigte die Auswertung bekannter, direkter Wirkstoff-Transporter-Beziehungen, dass die meisten sich auf die wenigen besser charakterisierten SLC-Familien bezogen, während nur wenige Ausnahmen die Beteiligung anderer SLC-Untergruppen beschrieben. Eine strukturierte und detaillierte Untersuchung eines möglichen Zusammenhangs zwischen der Familie der SLC-Transporter und der Wirkstoffverfügbarkeit stellt deshalb einen wichtigen Ansatz zur Klärung der Transportmechanismen sowie -voraussetzungen bedeutender therapeutischer Wirkstoffe dar.

In dieser Arbeit wird daher mittels einer Kombination bioinformatischer sowie experimenteller Methoden eine systematische Analyse der Beziehungen zwischen SLCs und Wirkstoffen unternommen. Gestützt auf den zurzeit umfangreichsten pharmakogenomischen Datensatz, welcher die Sensitivität von 1000 katalogisierten Krebszelllinien gegenüber 265 Wirkstoffmolekülen erfasst, haben wir basierend auf den molekularen Eigenschaften von SLC und ABC Transportern (Genexpression, Kopienzahlvariationen – CNVs –, Einzelnukleotid-Varianten – SNVs –) Vorhersagemodelle für die Wirkstoff-Sensitivität entwickelt. Des Weiteren

entwarfen wir ein CRISPR-basiertes genetisches Screening-Konzept mit zugehöriger Analyse-Pipeline, mittels dessen wir 60 unterschiedliche cytotoxische Substanzen hinsichtlich ihrer Abhängigkeit von Transportern experimentell testen konnten. Beide Ansätze erwiesen sich als erfolgreich in der Identifikation bereits bekannter sowie bislang noch nicht beschriebener Transporter-Wirkstoff Beziehungen, welche sowohl direkter als auch indirekter Natur waren und des Weiteren zum Teil experimentell validiert wurden.

Die hier vorgelegte Arbeit beschreibt die methodische, objektive Priorisierung potentiell pharmakologisch-relevanter Interaktionen und bringt neue Erkenntnisse hinsichtlich des Transportweges und der Pharmakodynamik mehrerer bedeutender therapeutischer Wirkstoffe. Dieses Verständnis ist von großem Belang für die Steigerung der Spezifität und Effizienz medikamentöser Therapien sowie den Übergang zu einer individualisierten Medizin.

Abbreviations

ABC	ATP-Binding Cassette
ACE	Angiotensin-Converting Enzyme
ADME	Administration, Distribution, Metabolism, Excretion
ANOVA	ANalysis Of VAriance
ATP	Adenosine TriPhosphate
AUC	Area Under the dose-response Curve
BBB	Blood-Brain Barrier
BCRP	Breast Cancer Resistance Protein
Cas	CRISPR-associated
CCL	Cancer Cell Line
CCLE	Cancer Cell Line Encyclopedia
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
CNS	Central Nervous System
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
CTRP	Cancer Therapeutics Response Portal
CV	Cross-Validation
dCas9	dead Cas9
DSB	Double Strand Break
ES	Enrichment Score
FACS	Fluorescence-Activated Cell Sorting
FDR	False Discovery Rate
GDSC	Genomics of Drug Sensitivity in Cancer
GLM	Generalized Linear Model
GPCR	G Protein-Coupled Receptor
GR	Growth Rate
GSEA	Gene Set Enrichment Analysis
HGNC	HUGO Gene Nomenclature Committee
HDR	Homology-Directed Repair
IC_{50}	half maximal Inhibitory Concentration

KF-CV	K-Fold Cross-Validation
KO	Knock-Out
LASSO	Least Absolute Shrinkage and Selection Operator
LDH	Lactate DeHydrogenase
LeuT	Leucine Transporter
LFC	Logarithm of the Fold Change
MAGECK	Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout
MANOVA	Multivariate ANalysis Of VAriance
MATE	Multidrug And Toxin Extrusion
MCA	Multicolor Competition Assay
MDR	MultiDrug Resistance
MFS	Major Facilitator Superfamily
ML	Maximum Likelihood
MoA	Mechanism of Action
MRP	Multidrug Resistance-associated Protein
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium
NB	Negative Binomial distribution
NBD	Nucleotide-Binding Domain
NCI	National Cancer Institute
NGS	Next-Generation Sequencing
NHEJ	Non-Homologous End Joining
NSAID	NonSteroidal Anti-Inflammatory Drug
OAT	Organic Anion Transporter
OATP	Organic Anion-Transporting Polypeptide
OCT	Organic Cation Transporter
PAM	Protospacer-Adjacent Motif
PCA	Principal Component Analysis
RNA-Seq	RNA Sequencing
RNAi	RNA interference
Ro5	Lipinski's Rule of 5
ROC	Receiver Operating Characteristic curve
RRA	Robust Ranking Aggregation
ScFv	Single-chain variable Fragment
(s)gRNA	(single) guide RNA

SLC	SoLute Carrier
SNV	Single Nucleotide Variant
SRB	SulfoRhodamine B
SSE	Sum of Squared Error
TALEN	Transcription Activator-Like Effector Nuclease
TM	TransMembrane
TMD	TransMembrane Domain
tracrRNA	Trans-activating CRISPR RNA

Acknowledgements

The completion of this doctoral thesis would not have been possible without the contribution and constant support of a number of persons whom I would like to acknowledge.

First of all, I would like to express my gratitude to my supervisor *Giulio Superti-Furga* for putting his trust in me since the very beginning and for offering me the opportunity of doing a PhD in his group. I highly appreciate his enthusiasm for science, his out-of-the-box thinking and his vision, but more importantly the support he provided me with and the patience and honesty he always had with me. I do not forget to also thank my former supervisor *Jacques Colinge* and his –now extinct– group for welcoming and introducing me to CeMM in the first place. I would also like to acknowledge *Julio Saez-Rodriguez*, *Mi Yang*, *Marc Brehme* and all members of their groups for kindly hosting me in Aachen and for establishing a fruitful collaboration.

I am extremely thankful to all the *GSF group members* –students, postdocs, technicians, staff– for their kindness, their positive attitude and their professionality, as well as for being able to create such an amazing lab atmosphere and for sharing so many good experiences together. In particular, I would like to thank *Ulrich Goldmann*, for being always willing to answer my bioinformatics questions and for critically reading this thesis, and especially to *Enrico Girardi*, for being such a patient, understanding, and professional colleague –and mentor–. I would not have been able to write this thesis and finish this PhD without his help and support.

I also want to acknowledge every single *CeMMie* for making of CeMM such a great and unique research institute and for all their scientific, technical and administrative support. And to *A.E.* for removing all barriers in the way of science –and easing so much the writing of this thesis.

Huge thanks go to the 2013 PhD student generation –*Máté, Paul, Michel, Thea, Aniko, Juliane, Bianca and Anna*– who have become true friends in the course of these years and have made of this an unforgettable experience.

Last but not least, I would like to thank all my *friends* in Pamplona, Geneva, Vienna and elsewhere, as well as my *family*, for always being there and always believing in me.

1 Introduction

The introduction of this thesis provides a broad overview of the main topics and methods covered by the three scientific articles presented in the results section, and is divided into three substantially different parts consisting of:

i) a review of the mechanism of drug transport across biological membranes, with a focus on the two main transporter superfamilies involved, and more in particular, the group of solute carriers (SLCs). This part serves as a common link among the three above-mentioned publications.

ii) a description of drug sensitivity predictive modeling using pharmacogenomics datasets of cancer cell line panels, with an emphasis on existing computational approaches and the method used in the second publication.

iii) a description of the CRISPR-Cas9 method for pooled knock-out genetic screening including existing computational methods most commonly used for its analysis and the strategy applied in the third and last article of this thesis.

1.1 Drug Transport and Drug Transporters

Most drugs need to cross at least one biological membrane in order to reach their target and exert their action. Transport is therefore a key process in drug pharmacokinetics, as it directly controls the extent and rate at which compounds enter the organism, determines their differential distribution across tissues and organs, and mediates their elimination from the body. Understanding the specific mechanisms by which drugs enter and exit cells is thus essential in order to increase the probability of therapeutic success while reducing the chance of adverse reactions, and their study constitutes an integral part of every drug discovery program (Goodman et al, 2011).

1.1.1 Drug transport across biological membranes

The passage of solutes through cellular membranes is regulated by different mechanisms that can be classified according to thermodynamic and mechanistic criteria (**Figure 1**). *Passive* processes are equilibrative, allowing the movement of solutes in the direction of their electrochemical gradient ("downhill"), and therefore do not have any extra energy requirements. Although the term "passive" does not *per se* imply a specific mechanism, it is often understood as *simple* or *lipoidal* diffusion (i.e. directly through the lipid bilayer), and referred as *facilitated* diffusion when it uses the help of a carrier. In contrast, *active* transport is concentrative, moving solutes against their gradient ("uphill"), for which it strictly requires a source of metabolic energy and the obligatory participation of specific proteins. Depending on the energy source, this transport type is further subdivided into *primary active*, when energy is directly obtained from the hydrolysis of ATP, and *secondary active*, when it is provided by the electrochemical potential of another solute that is co-transported with the substrate, either in the same (i.e. symport) or opposite (i.e. antiport, exchange) direction (Alberts et al, 2014; Kell & Oliver, 2014).

The standard view of cellular membrane architecture depicted by the "fluid mosaic" model considered proteins being dispersed and at low concentration within the lipid bilayer (Singer & Nicolson, 1972). However, it is nowadays well accepted that these are present at a much higher density, compartmentalizing membranes into functional domains and giving them a more "patchy" nature (Engelman, 2005; Goñi, 2014; Nicolson, 2014). In relation to this, there is still an

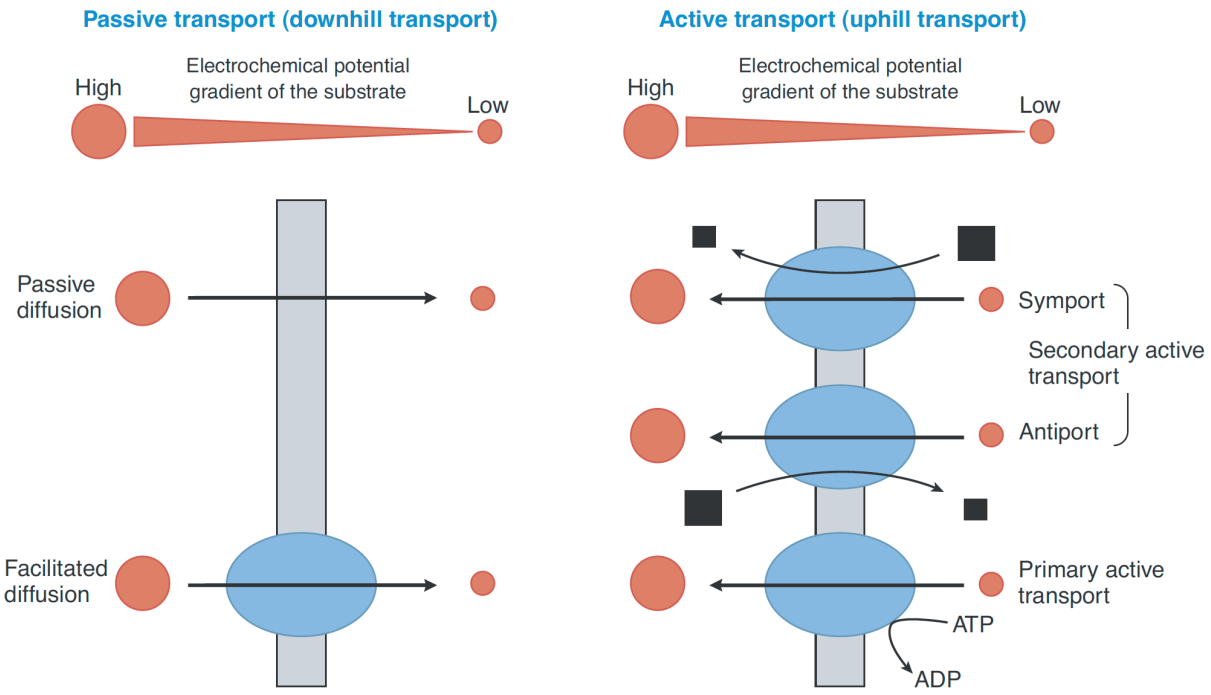


Figure 1 Classification of membrane transport mechanisms. Transport processes are classified into passive or active according to their energy requirements, and further subdivided based on the participation of specific proteins, energy source and transport directionality. Red circles and black squares represent the substrate and the ion that supplies the driving force for transport, respectively; sizes are proportional to the concentration of the solutes; arrows display the direction of flux; and blue ovals represent transporter proteins.

Reprinted from (Goodman et al, 2011) with permission from The McGraw-Hill Companies, Inc.

ongoing debate on the relative contribution of carrier proteins to the transport of drugs. Passive diffusion through the lipid bilayer has generally been considered the dominant process by which chemical compounds enter and exit cells, which is highly dependent on drug features such as molecular size, degree of ionization or lipid-water partition. Indeed, Lipinski's "rule of 5" (Ro5), a frequently used rule of thumb to evaluate *druglikeness* of orally administered compounds, bases absorption on size and lipophilicity, and considers drugs that act as substrates of transporters exceptions to the rule (Lipinski et al, 2001). However, some studies have pointed to 20-30% of FDA-approved small-molecule drugs not fulfilling Ro5 (Overington et al, 2006), and there is a growing number of known cases of drugs whose transport is mediated by specific transporters (Fets et al, 2018; Winter et al, 2014; Yu et al, 2018; Zhao et al, 2011). These and other considerations have prompted some authors to question the importance of lipoidal passive

diffusion in favor of carrier-mediated drug transport (Dobson & Kell, 2008; Kell, 2015; Kell et al, 2011) while others defend a more even coexistence of both mechanisms (Di et al, 2012; Matsson et al, 2015; Smith et al, 2014; Sugano et al, 2010).

Whether the rule or the exception in mediating the uptake and efflux of xenobiotics, transporters are increasingly recognized, together with drug-metabolizing enzymes (DMEs) such as cytochromes and transferases, as the main determinants of drug disposition (Zhou et al, 2017). They are indeed directly involved in the selective absorption, distribution and excretion of chemical compounds, often representing the rate-limiting step of these processes (Giacomini & Huang, 2013). In particular, transporters allow tissue-specific drug targeting (Zhou et al, 2015) and also exert an essential role in the detoxification and protection of vital tissues against xenobiotics (Iqbal et al, 2012; Leslie et al, 2005; Mahringer & Fricker, 2016). Such a role is highlighted by their characteristic expression in important pharmacological barriers, such as the membranes of intestinal cells, hepatocytes and proximal tubules of the kidney as well as the brain-blood (BBB), blood-testes and blood-placenta barriers (César-Razquin et al, 2015; O'Hagan et al, 2018). Additionally, they are mediators of drug-drug and nutrient-drug interactions (Bailey, 2010; König et al, 2013; Zhang et al, 2014) as well as drug toxicity (DeGorter et al, 2012). As such, regulatory guidelines are being constantly published and updated in order to guide the drug development process (International Transporter Consortium et al, 2010; Zamek-Gliszczynski et al, 2018). Finally, they play an important role in the development of resistance to therapeutic agents such as anticancer drugs (Robey et al, 2018; Szakács et al, 2006) or antivirals (Ibarra & Pfeiffer, 2009; Imaoka et al, 2007).

1.1.2 Classes of drug transporters

It has been estimated that ~10% of human genes encode for transport-related functions (Hediger et al, 2013). Membrane transport proteins can be roughly classified into two main types, *channels* and *carriers* (often simply referred as *transporters*), whose main difference resides in their mechanism of transport. Channels form hydrophilic pores across the lipid bilayer that, in the open state, permit the rapid diffusion of specific solutes (mostly inorganic ions of a certain size and charge) at a rate dependent on the channel conductance. In contrast, carrier proteins interact directly with their substrates through specific binding sites and experience a

series of conformational changes that allow the transfer of the bound cargo to the opposite side of the membrane, therefore presenting a fixed stoichiometry of substrate movement per translocation cycle (Alberts et al, 2014). Drug transporters belong to the latter class, and include two large superfamilies: ABC (ATP-binding cassette) and SLC (Solute carrier) transporters.

ATP-binding cassette transporters (ABCs)

ABCs are mostly primary active transporters that use the energy of the hydrolysis of ATP to translocate specific substrates across biological membranes. The main common characteristic of these proteins is the presence of a highly conserved *ATP-binding cassette* domain, also called *nucleotide-binding* domain (NBD), which gives the family its name. ABCs can be found in all domains of life, but while in prokaryotes they mediate both uptake and efflux, in eukaryotes they are only known to act as exporters that move compounds out of the cytoplasm. Human ABCs are present in the plasma as well as in intracellular membranes, and are able to pump out a plethora of substrates, ranging from metal ions to peptides, but more typically lipophilic molecules, playing a crucial role in the transport of lipids, fatty acids and cholesterol (Borst & Elferink, 2002; Rees et al, 2009). Consequently, mutations in more than half of human ABC genes have been linked to diseases, often related to defects in lipid homeostasis (Tarling et al, 2013).

The basic core unit of a *full* ABC is made of two highly conserved cytoplasmic NBDs, which are in charge of binding and hydrolyzing ATP, and two variable *transmembrane* domains (TMD), typically containing six α-helices each, which take care of recognizing and translocating the substrate (ter Beek et al, 2014) (**Figure 2**). Some ABCs exist as *half*-transporters, only containing one domain of each type (e.g. the ABCG family), but since two NBD are required for ATP hydrolysis, they need to at least dimerize as either homo- or heterodimers in order to be functional (Dezi et al, 2010). They can also present additional domains (e.g. some members of the ABCC family) as well as lack some of the core ones (Biemans-Oldehinkel et al, 2006). The mechanism of transport of ABCs is generally described using the "alternating access" model, in which conformational changes alternately expose the substrate binding site to the extracellular (*outward facing* conformation) and cytoplasmic (*inward facing* conformation) sides of the membrane, passing through other intermediate states such as an *occluded* conformation

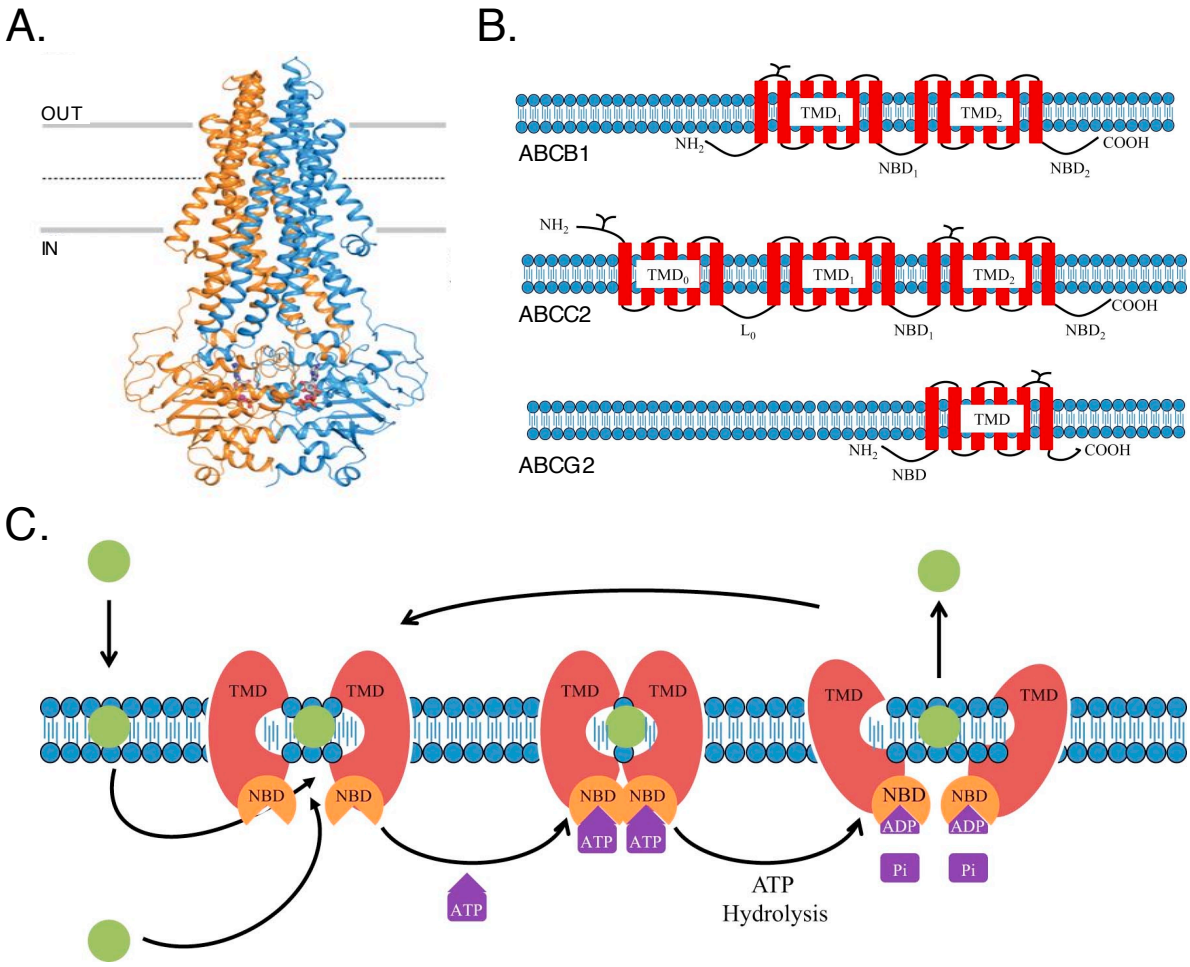


Figure 2 Structures and transport mechanisms of ABCs. A. Structure of ABCB1 in complex with ATP. The N- and C-terminal halves, containing one TMD and one NBD each, are colored in orange and blue, respectively. ATP is shown in ball-and-stick format (gray: C; red: O; blue: N; orange: P), and Mg^{2+} is shown as a magenta sphere. **B.** Secondary structure models of ABCB1 (P-gp), ABCC2 (MRP2), and ABCG2 (BCRP). Transmembrane domains (TMD), nucleotide-binding domains (NBD) and loop 0 (L_0) are indicated. **C.** Transport mechanism of ABC transporters. ABCs are energy-dependent (primary active) transporters. A conformational change upon substrate binding and ATP hydrolysis drives the translocation of the substrate.

A was reprinted from (Kim & Chen, 2018) with permission from Science. B and C were reprinted from (Chen et al, 2016) with permission from Elsevier.

where the binding site is not accessible from either side (Jardetzky, 1966). Nevertheless, due to the structural diversity among the members of the ABC superfamily, one-size-fits-all models might not describe accurately enough the actual transport mechanisms (Locher, 2016).

There are currently 51 human ABC genes listed by the HGNC, including three pseudogenes, classified phylogenetically into 7 different subfamilies named A to G (Dean et al, 2001; Gray et al, 2016; Vasiliou et al, 2009). However, not all of them are actually transporters. For instance, all members of the ABCE (one gene) and ABCF (three genes) subfamilies lack TM domains and have been implicated in functions such as viral infection (Dooher et al, 2007), ribosomal recycling (Pisarev et al, 2010) and regulation of translation (Mancera-Martínez et al, 2017; Paytubi et al, 2009). Another case is ABCC7 (or CFTR, for *Cystic Fibrosis Transmembrane Conductance Regulator*), which in spite of sharing the full domain structure of ABCs, acts as an ATP-gated chloride ion-channel whose mutation causes cystic fibrosis (Gadsby et al, 2006). Similarly, ABCC8 and ABCC9 (SUR1/2, respectively) serve as potassium channel regulators (Aittoniemi et al, 2009; Bryan et al, 2007).

Several members of the ABC family are well-known mediators of drug ADME, and due to their ability to export a large number of anticancer compounds, they often appear associated with *acquired multidrug resistance* (MDR) in cancer (Szakács et al, 2006). Most drug-transporting ABCs belong to three families:

— ***ABCB subfamily***. ABCB1, also known as *P-glycoprotein* (Pgp) or *Multidrug Resistance Protein 1* (MDR1), and described as the "double-edged sword" for its dual role in protection against xenobiotics and cause of MDR, is probably the most studied transporter (Gottesman & Pastan, 1988). It is mainly expressed in liver, kidney, small intestine and brain, playing an essential protective role at the blood-brain barrier (BBB). The main characteristic of Pgp resides in its unusual broad polyspecificity. Although it has a preference for large neutral or weakly basic hydrophobic substrates, often containing aromatic rings and charged tertiary amino groups, it can also inefficiently handle anionic molecules. It is therefore able to transport hundreds of structurally dissimilar compounds, ranging from 200 to greater than 1,000 Da in molecular weight, many of which are of great clinical importance, including anticancer drugs, antibiotics, antivirals, analgesics, and immunosuppressives (Borst & Elferink, 2002; Sharom, 2008). Indeed, the lack of highly conserved recognition elements makes it difficult to predict ABCB1 substrates (Chen et al, 2012). Testing for susceptibility to transport by Pgp is since long common practice in drug development programs (International Transporter Consortium et al, 2010) and inhibitors of Pgp (e.g. excipients like tocopherol) are

used in the clinic in order to enhance the bioavailability of drugs that are ABCB1 substrates (Hodges et al, 2011). Two other members of this family, ABCB4 (MDR3) and ABCB11 ("sister of Pgp", BSEP), are also known to mediate drug transport, but their expression is restricted to the liver and they present a much more limited substrate specificity as well as lower transport efficiency (Hirano et al, 2005; Morita & Terada, 2014).

— ***ABCC subfamily***. Most members of this family, also called *Multidrug resistance-associated proteins* or MRPs, mediate the transport of anionic hydrophilic compounds. This includes glutathione (GSH), sulfate and glucuronate conjugates of hydrophobic drugs that result from the action of metabolizing enzymes during phase II metabolism. The different members of the family differ in their substrate specificity, tissue expression, cellular location and structure, which includes in some cases an additional N-terminal TM domain. ABCC1 (MRP1) is an ubiquitous and versatile transporter, which also confers resistance to several neutral/basic hydrophobic compounds that are substrates of ABCB1. ABCC2 (MRP2), probably the most relevant pharmacologically, overlaps significantly with ABCC1 in its substrate specificity, but has a more restricted tissue distribution (mainly to liver, intestine and kidney), and it localizes to the apical membrane of epithelial cells, whereas ABCC1 and most of the other members are basolateral. ABCC3 (MRP3) has the highest sequence similarity to ABCC1 (58%) but transports fewer compounds and shows a preference for glucuronides over GSH. Other members, such as ABCC4 (MRP4) and ABCC5 (MRP5), are in addition able to transport cyclic nucleotides and nucleotide analogs (e.g. 6-mercaptopurine or 5-thioguanine) at a low rate (Borst & Elferink, 2002; Kruh & Belinsky, 2003; Szakács et al, 2006).

— ***ABCG subfamily***. The most important member of this family regarding drug disposition is ABCG2, which was originally given three other names based on its main reported characteristics: overexpression in breast cancer cell lines (*Breast Cancer Resistance Protein*, BCRP), enrichment in placental tissue (*Placenta-specific ABC*, ABCP) and mitoxantrone transport (*Mitoxantrone Resistance-associated Protein*, MXR). ABCG2 is a high-capacity half transporter that functions as a homodimer and is expressed in several tissues, mainly in placenta, liver, intestine, endothelium and mammary tissue. It has a wide substrate specificity, being able to transport large molecules of amphipathic character,

both positively and negatively charged, including many chemotherapy agents. Many of its substrates overlap with ABCB1, but in contrast to this protein and similarly to MRPs, it can also efficiently mediate the efflux of anionic drug conjugates (Robey et al, 2009; Sarkadi et al, 2006).

Solute carriers (SLCs)

SLCs are the second-largest group of membrane proteins in the human genome, after G protein-coupled receptors (GPCRs). They constitute a heterogeneous collection of transporters whose common feature is that they do not rely directly on the hydrolysis of ATP in order to exert their function, thus acting as either facilitative or secondary active transporters (symporters or antiporters). The inclusion of a protein within the SLC group is therefore based on functional criteria, and not homology (Höglund et al, 2011). SLCs are present in the plasma membrane as well as in intracellular organelles (e.g. the mitochondrial SLC25 family) (Palmieri, 2013) and, although they are mainly involved in the uptake of solutes into the cytoplasm, a few of them are well-known exporters (e.g. SLC47 family) (Motohashi & Inui, 2013). They are able to transport an enormous variety of solutes including charged and uncharged organic molecules (e.g. sugars, nucleotides, amino acids, peptides, fatty acids), inorganic ions or even the gas ammonia (SLC42 family) (Weiner & Verlander, 2010) (Hediger et al, 2004).

The total number of human SLCs has been constantly rising in the last years (Perland & Fredriksson, 2017). The most recent count involves more than 450 genes, including ~20 pseudogenes, classified into 66 different numbered families (Gray et al, 2016) (**Figure 3**). SLC families are defined using a cut-off of 20-25% amino acid sequence similarity among its members, which are usually identified by a unique root symbol that includes the family number (Hediger et al, 2013). However, exceptions to this rule exist. For instance, SLC51 family contains two members that do not share sequence similarity but encode two subunits of the same transporter (Ballatori et al, 2013), SLC21 family changed its root symbol to SLCO in order to accommodate an already existing species-independent classification system (Hagenbuch & Meier, 2004) and other members such as UCPs (SLC25A7 to 9), Rh glycoproteins (SLC42A1 to 3) or FLVCRs (SLC49A1 and 2) have kept their original gene symbols. Moreover, a set of putative transporters commonly known as *atypical SLCs* are not entirely classified yet into any SLC family but are likely to be included once they are further characterized (Perland et

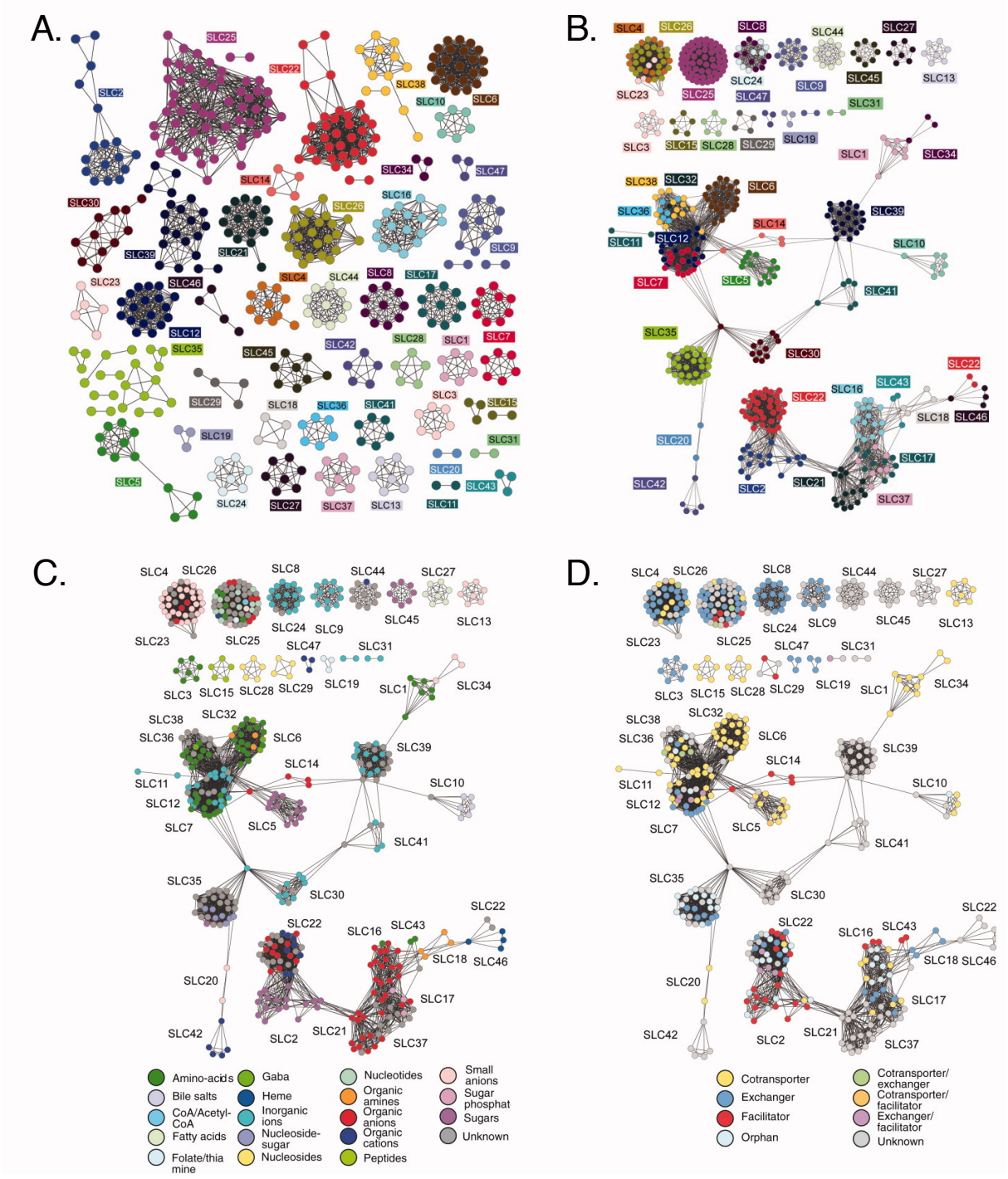


Figure 3 **Classification of SLCs.** Similarity networks for certain SLC families. **A.** Links represent pairwise alignments with sequence identity $\geq$25% and E-value $<$1. Colors indicate different SLC families. **B.** Links represent pairwise alignments with sequence identity $\geq$10% and E-value $<$1. Colors as in A. **C.** Same as B., but colors represent the prototypical substrates of the transporters. **D.** Same as B., but colors represent transport mode.

Reprinted from (Schlessinger et al, 2010) with permission from John Wiley and Sons.

al, 2017). Other classification systems for SLCs have also been proposed that are based on function, structure, transport mechanism and phylogenetic relatedness (Finn et al, 2016; Saier et al, 2016; Schlessinger et al, 2010). In this sense, phylogenetic studies have shown that at least some SLC families are evolutionarily related, forming four distinct phylogenetic clusters (Fredriksson et al, 2008). Similarly, these analyses have reported that SLCs present the richest evolutionary history of all membrane protein groups, with more than 50% of human families being found in Prokaryotes and Archea (Höglund et al, 2011). Also, not all characterized SLCs are actual transporters. Some do not have transport capability on their own and need to interact with other members in order to form functional complexes, such as SLC3 and SLC7 or SLC51A and B. Others are even able to exert functions unrelated to transport, such as signal sensing and transduction (Rebsamen et al, 2015).

SLCs are structurally very diverse. They are usually made of 10-14 transmembrane (TM) domains and cover a number of different folds, two of which are significantly represented: the *Major Facilitator Superfamily* (MFS) fold and the *Leucine Transporter-like (LeuT)* fold (Colas et al, 2016) (**Figure 4a,b**). MFS is the most common, predicted to be present in at least 17 SLC families (e.g. SLC2, SLC15, SLCO, SLC22). It comprises 12 TM helices arranged into two 6-TM homologous domains of "3+3" inverted repeats (the *N*- and *C-terminal* domains) connected by a cytoplasmic loop (Yan, 2015). In contrast, the LeuT fold (e.g. SLC5, SLC6) has a core structure of 10 TM helices made of two 5-TM inverted pseudorepeats, where the first two TMs of each repeat form a four-helix *core* domain and the third and fourth TMs constitute a *scaffold* domain surrounding it (Yamashita et al, 2005). Other members present less common folds involving inverted repeats, hairpin domains or even channel-like structures (e.g. SLC14, SLC42) (Bai et al, 2017). At least three types of alternating access mechanisms have been proposed for SLC-mediated transport (**Figure 4c**): *i) Rocker-switch*, which describes the oscillation of the N and C-terminal domains of the transporter (e.g. SLCs with MFS fold); *ii) Gated-pore*, in which two gates control the sequential access to the substrate binding site from either side of the membrane (e.g. for LeuT folds), and *iii) Elevator*, which depicts the movement of the domain containing the binding site (*transport domain*) along the axis perpendicular to the membrane (e.g. SLC1 family) (Colas et al, 2016).

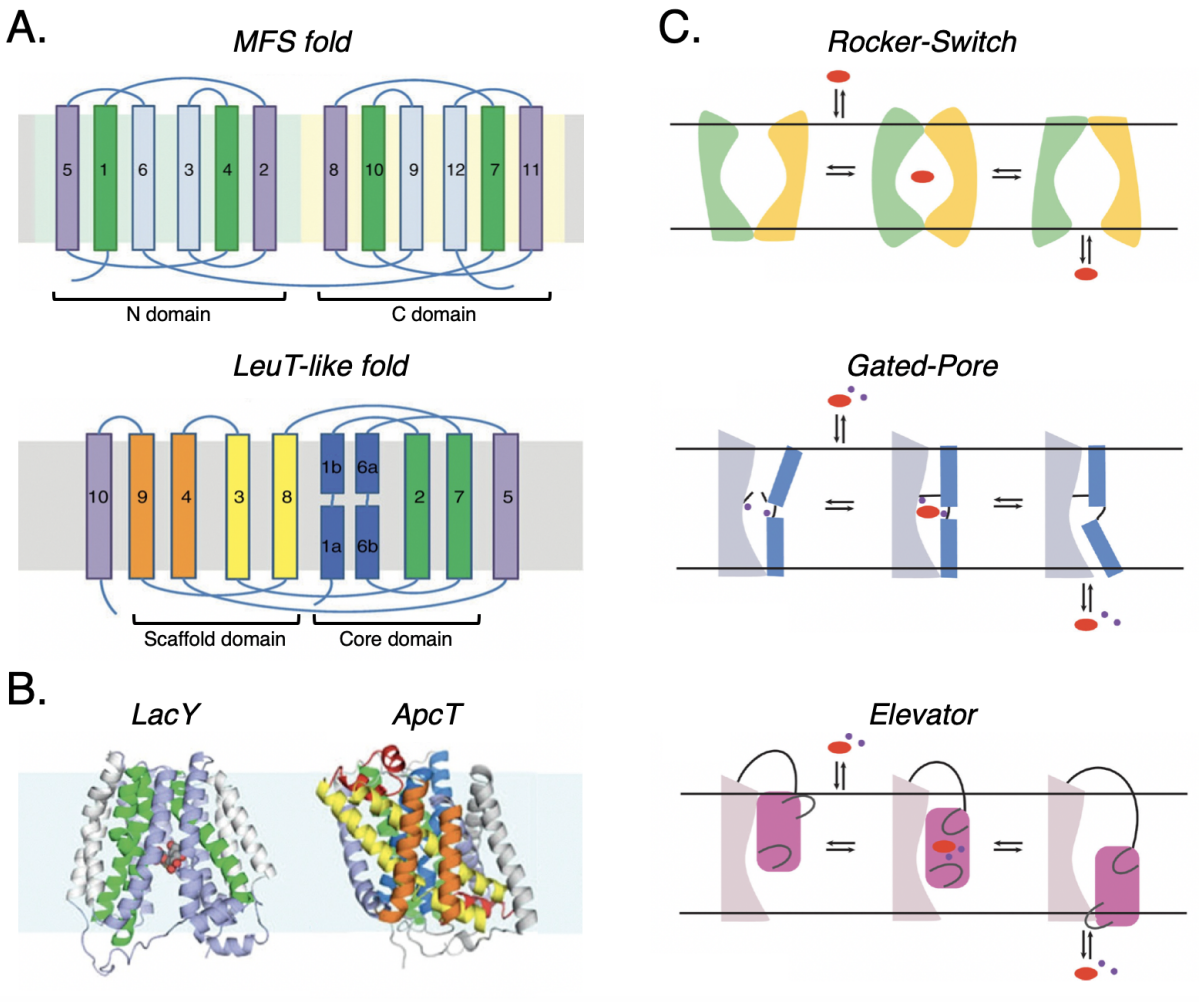


Figure 4 Structures and transport mechanisms of SLCs. A. Topology diagram of the Major Facilitator Superfamily (MFS) and Leucine Transporter-like (LeuT) folds. Transporters are positioned with their periplasmic side at the top and cytoplasmic side at the bottom. Corresponding TMs in each inverted repeat are colored the same. **B.** Representative structures of transporters with MFS and LeuT-like folds (from *E.coli*). Color scheme as in A. **C.** Transport mechanism models. Outward-open, occluded and inward-open states are represented for each model.

A and B were adapted from (Yan, 2013) with permission from Elsevier. C was adapted from (Colas et al, 2016) with permission from The Royal Society of Chemistry.

Importantly, structure and transport specificity or function do not form a one-to-one relationship. Members of the same fold or family can sometimes mediate the transport of chemically dissimilar solutes (e.g. SLC22 family) and very similar substrates are often transported by structurally different families (e.g. glucose and SLC2 – MFS fold –, SLC5 – LeuT fold –, and SLC50 families). Such within-family varied degree of specificity and inter-family

functional redundancy are actually two of the most defining characteristics of SLCs (Colas et al, 2016).

The ability of SLCs to transport all major building blocks of the cell makes them key mediators of cell physiology and homeostasis (**Figure 5**). Indeed, defects in SLC function are the cause or have been associated with numerous human diseases, from monogenic to multifactorial pathologies such as neurological and mental disorders, inflammatory bowel disease, type II diabetes or cancer (Lin et al, 2015; Zhang et al, 2018). Several SLCs are currently the target of approved drugs and drugs in development, and many more seem to be potentially druggable (Rask-Andersen et al, 2013; Rask-Andersen et al, 2014). As an illustration, serotonin transporter SLC6A4 (SERT) is the target of the most commonly used class of antidepressants worldwide, which includes the well-known drug fluoxetine (Prozac) (Andersen et al, 2014), and inhibitors of the glucose transporter SLC5A2 (SGLT2) are since recently used as a treatment of type II diabetes (Scheen, 2015). However, in spite of their central role in cellular metabolism and their relevance for pharmacology, SLCs are the most neglected human gene group as a whole, with the vast majority of its members still poorly characterized (César-Razquin et al, 2015).

Most members of the SLC group that are well-known mediators of drug disposition belong to four families (DeGorter et al, 2012; Estudante et al, 2016; International Transporter Consortium et al, 2010):

— **SLC22 family** *of organic anion/cation/zwitterion transporters*. Probably the best-studied SLC family regarding drug disposition, it comprises more than 20 members grouped into two major clades: *OCT* (organic cation transporter) and *OAT* (organic anion transporter). While some of them present a broad substrate spectrum (e.g. OAT1, OAT3, OCT1, OCT2), others have more restricted specificity (e.g. OAT6, URAT1, OATPG, OCTN2, OCTN1). OCTs are mainly facilitative transporters that mediate the uptake of small positively charged molecules, including platinum-containing anticancer agents (e.g. oxaliplatin), antidiabetic metformin, or the antiulcer and acid reducer cimetidine. The most relevant members include SLC22A1 (OCT1) and SLC22A2 (OCT2), which transport similar substrates and are expressed in liver and kidney, respectively, and SLC22A3 (OCT3), which has a wider tissue expression including the central nervous system (CNS). OATs, in contrast, are exchangers that use

endogenous dicarboxylic acids (e.g. alpha-ketoglutarate) to move small organic anions against their concentration gradient, such as antibiotics (e.g. tetracycline), antivirals (e.g. acyclovir), antihypertensives (e.g. olmesartan, captorpril), NSAIDs (e.g. ibuprofen) or the chemotherapy agent methotrexate. SLC22A6 (OAT1), expressed predominantly in the basolateral membrane of proximal renal tubules, and SLC22A8 (OAT3), present ubiquitously in the kidney and the choroid plexus, are the most important members for drug disposition and have a broad and overlapping substrate specificity, although SLC22A8 shows a greater preference for slightly cationic as well as larger and more complex molecules (Koepsell, 2013; Liu et al, 2016; Nigam, 2018).

— **SLCO (SLC21) family** *of organic anion transporters*. Also called *organic anion-transporting polypeptides* (OATPs), they mediate the sodium-independent uptake of a structurally diverse range of amphipathic organic molecules, probably via anion exchange with solutes such as bicarbonate, although their exact transport mechanism is not yet completely clear. Drugs transported by OATPs are statins, antidiabetics, chemotherapy agents (e.g. methotrexate, irinotecan), antibiotics (e.g. rifampicin) or antihypertensives (e.g. olmesartan), among other. The two best-studied transporters of this family, SLCO1B1 (OATP1B1) and SLCO1B3 (OATP1B3), are both expressed in the basolateral membrane of hepatocytes and transport similar substrates, although SLCO1B3 shows a unique transport specificity towards taxanes (e.g. paclitaxel, docetaxel), cardiac glycosides (e.g. digitoxin, ouabain) and small peptides. Other relevant members are SLCO2B1 (OATP2B1) and SLCO1A2 (OATP1A2), which have a broader expression pattern and play a key role at the blood-brain barrier (BBB) (Hagenbuch & Gui, 2008; Leuthold et al, 2009).

— **SLC47 family** *of multidrug and toxin extrusion transporters (MATE)*. MATEs are obligatory exchangers that use the electrochemical gradient of H^+ in order to flush numerous compounds out of cells and, similarly to ABCs, often confer MDR. There are only two human members in this family: SLC47A1 (MATE1), predominantly expressed in kidney and liver, and SLC47A2 (MATE2), more ubiquitously expressed but with a specific kidney isoform, MATE2-K. They are multiselective exporters whose preferred substrates are of cationic nature and tend to overlap with those of OCTs (e.g. metformin, cimetidine, oxaliplatin), although MATEs are also able of transporting several zwitterionic (e.g. antibiotics cephalexin

and cephradine, only MATE1) and anionic (e.g. acyclovir, ganciclovir) compounds. Importantly, they operate in concert with OCTs in the transepithelial vectorial transport of many solutes, where OCTs mediate basolateral uptake and MATEs support efflux through the apical membrane, often co-localized with members of the ABC family (Motohashi & Inui, 2013; Staud et al, 2013).

— **SLC15 family** *of proton-oligopeptide cotransporters*. SLC15 members are symporters that mediate the cellular uptake of a wide variety of di- and tripeptides of differing charge, hydrophobicity and size using the driving force of the inwardly directed H^+ gradient. Substrates include a number of peptide-like drugs such as aminocephalosporins (e.g. cefadroxil), ACE inhibitor antihypertensives (e.g. enalopril, captopril) or the anticancer agent bestatin, as well as *prodrugs* that, by coupling an oligopeptide to the actual compound structure, can hijack on these transporters and be more efficiently absorbed (e.g. valacyclovir). This makes SLC15 transporters very good targets for drug delivery. The two best-characterized members of this family are SLC15A1 (PEPT1) and SLC15A2 (PEPT2), which present very similar substrates. However, while SLC15A1 is a low-affinity and high-capacity intestinal transporter that plays an important role in the absorption of small peptides from dietary protein digestion and is also present in kidney and liver, SLC15A2 has high-affinity but low-capacity and is more ubiquitously expressed, especially at the brush border of renal proximal tubules, choroid plexus, lungs and skin (Rubio-Aliaga & Daniel, 2002; Smith et al, 2013).

Several other cases of drug disposition mediated by SLCs that do not belong to these four families also exist (Estudante et al, 2016). Often they involve transporters with narrower specificities and compounds with very similar structures and chemical properties to the actual endogenous substrates, in agreement with the concept of "metabolite-likeness" (Dobson et al, 2009). As an illustration, members of the concentrative and equilibrative nucleoside transporter SLC28 and SLC29 families mediate the uptake of nucleoside analogs (e.g. gemcitabine, zidovudine) (Young, 2016) and anti-folates such as methotrexate can be transported by the two folate transporters SLC19A1 and SLC46A1 (Zhao & Goldman, 2013). Other cases might involve *orphan* transporters i.e. carriers for which no endogenous substrates are known. For instance, SLC35F2 was identified as the specific importer of the anti-cancer drug

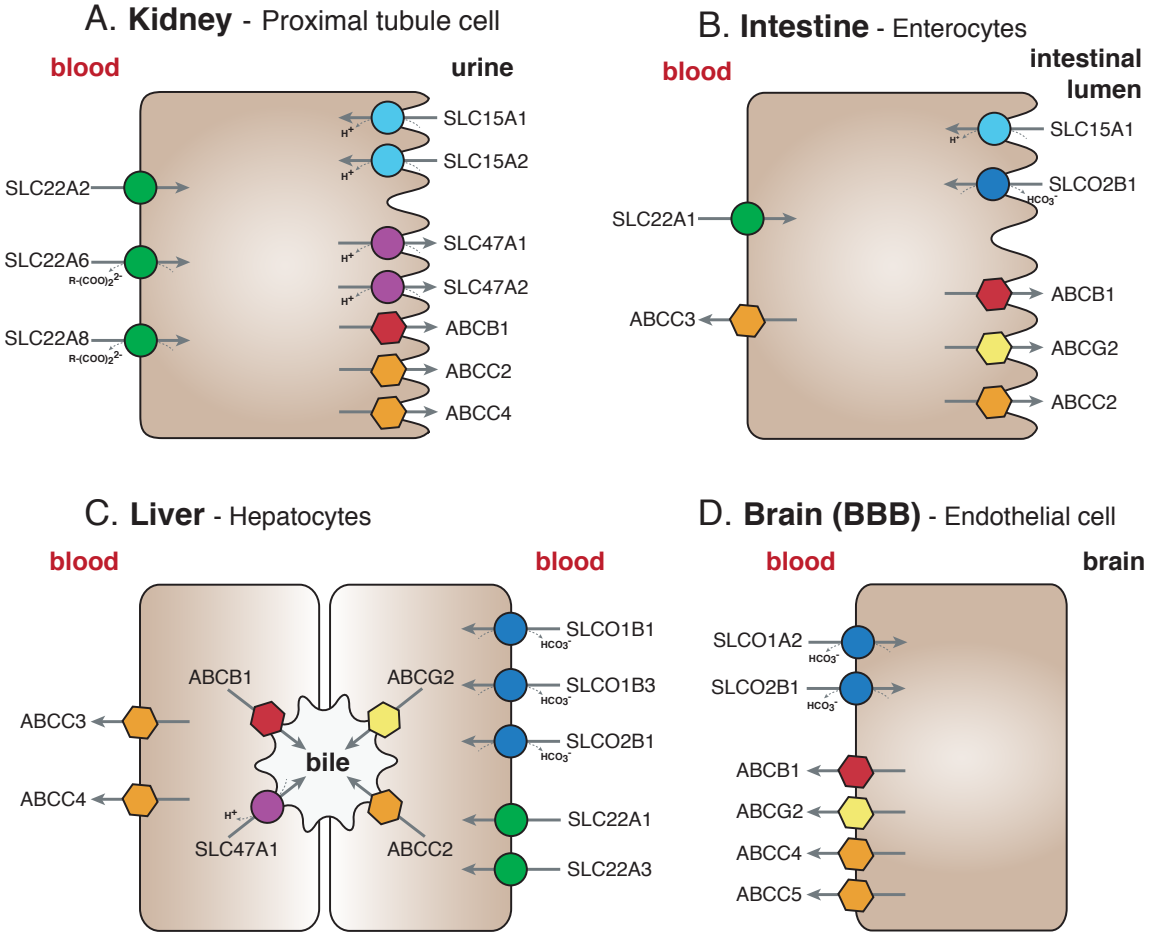


Figure 5 Coordinated action of SLC and ABC drug transporters in selected tissues. Colors depict individual families: ABCB (red), ABCC (orange), ABCG (yellow), SLC15 (light blue), SLCO (dark blue), SLC22 (green), SLC47 (purple). For secondary active transporters the ion that provides the energy for transport is indicated.

Based on (DeGorter et al, 2012; International Transporter Consortium et al, 2010)

YM155/sepantronium bromide (Winter et al, 2014) and SLC37A3 was recently reported to mediate the release of nitrogen-containing biphospohonates from lysosomes into the cytosol (Yu et al, 2018).

Due to the understudied nature of Solute carriers and the lack of information regarding the transport mechanism of many compounds, it is to expect that many new drug transport cases involving SLCs will be described in the future (César-Razquin et al, 2018).

1.2 Drug Sensitivity Prediction using Cancer Cell Line Pharmacogenomics Datasets

Cultured cancer cell lines (CCLs) are the most widely used model systems to study the molecular basis of drug activity, and constitute extremely valuable tools for drug discovery. The advent of high-throughput technologies has permitted the systematic characterization of thousands of CCLs at the molecular level, including gene expression, genomic variations and DNA methylation, among other. Simultaneously, profiling cancer cell line panels against large collections of small molecules has become a valuable method to associate drug response with baseline molecular descriptors. Predictive models of drug sensitivity built with these pharmacogenomics datasets have allowed the discovery of genetic markers of drug sensitivity, the identification of the mechanism of action (MoA) of compounds as well as the repurposing of drugs, and they ultimately aim at identifying the most suitable treatments for individual patients based on their genomic make-up, this is, *personalized* -or precision- medicine (Weinstein et al, 1992).

1.2.1 Large-scale pharmacogenomics datasets in cancer cell line panels

In the last two decades, a number of consortia have launched large-scale pharmacogenomics projects that involved the characterization of hundreds of cancer cell lines (Weinstein, 2012) (**Table 1**). The *Developmental Therapeutics Program* (DTP) of the *United States National Cancer Institute* (NCI) pioneered these initiatives in 1990 with the *NCI-60 Human Tumor Cell Lines Screen*, a collection of 60 (now 59) cell lines covering 9 different cancers aimed at screening thousands of small molecules for anticancer activity (Shoemaker, 2006). The NCI-60 panel contains the most extensively characterized cell lines, profiled at DNA, RNA, protein and even metabolite levels (Jain et al, 2012), and has the largest compound library, with more than 100,000 synthetic and natural products tested to date. However, an important downside of this resource resides in its limited cell line coverage.

More recent multi-omics drug sensitivity screens are the *Cancer Cell Line Encyclopedia* (CCLE) (Barretina et al, 2012) and the *Cancer Therapeutics Response Portal* (CTRP) (Basu et al, 2013; Seashore-Ludlow et al, 2015), two related initiatives led by the Broad Institute, as well

	NCI-60	CCLE	CTRP	GDSC
# cell lines	60	1457	860	1001
# compounds	>120k	24	481	265
main omics datasets	SNVs, CNVs, Meth, Gex, Prot, Metab	SNVs, CNVs, Meth, Gex	SNVs, CNVs, Gex	SNVs, CNVs, Meth, Gex
# cancer tissues	9	24	24	29
assay type	metabolic (sulforhodamine B)	metabolic (ATP luciferase)	metabolic (ATP luciferase)	DNA dye or metabolic (Resazurin or ATP luciferase)
reference	(Shoemaker, 2006)	(Barretina et al, 2012)	(Basu et al, 2013)	(Yang et al, 2013)

Table 1 Main publicly available large-scale pharmacogenomic resources. SNV: single nucleotide variants, CNV: copy number variations, Meth: methylation, Gex: gene expression, Prot: proteomics, Metab: metabolomics.

as the *Genomics of Drug Sensitivity in Cancer* (GDSC) (Garnett et al, 2012; Iorio et al, 2016) (**Figure 6**), a collaborative effort of the Wellcome Trust Sanger Institute and the Massachusetts General Hospital Cancer Center. They all offer larger numbers of cell lines with a broader tissue coverage than the NCI-60 panel, annotated with *Copy Number Variations* (CNVs), *Single Nucleotide Variants* (SNVs) and gene expression data, and tested against a few hundred of compounds (see **Table 1**).

1.2.2 Small-molecule in-vitro screening

High-throughput drug sensitivity screens consist in the assessment of the cytotoxic effect of collections of compounds across a wide range of concentrations on panels of cell lines in a miniaturized and automated manner. In these screens, cytotoxicity is evaluated by measuring reporter signals of cellular features whose change is associated with cell death or proliferation, such as cell number, metabolic state, or membrane integrity, under two main assumptions: *i)* the signals are correlated to cytotoxicity, and *ii)* the relationship between cell viability and compound concentration is monotonic (Cortes-Ciriano et al, 2016).

At least three main classes of cytotoxicity assays can be distinguished. *Colony formation* or *colonogenic assays* evaluate the capacity of treated cells to form colonies, and are more suited

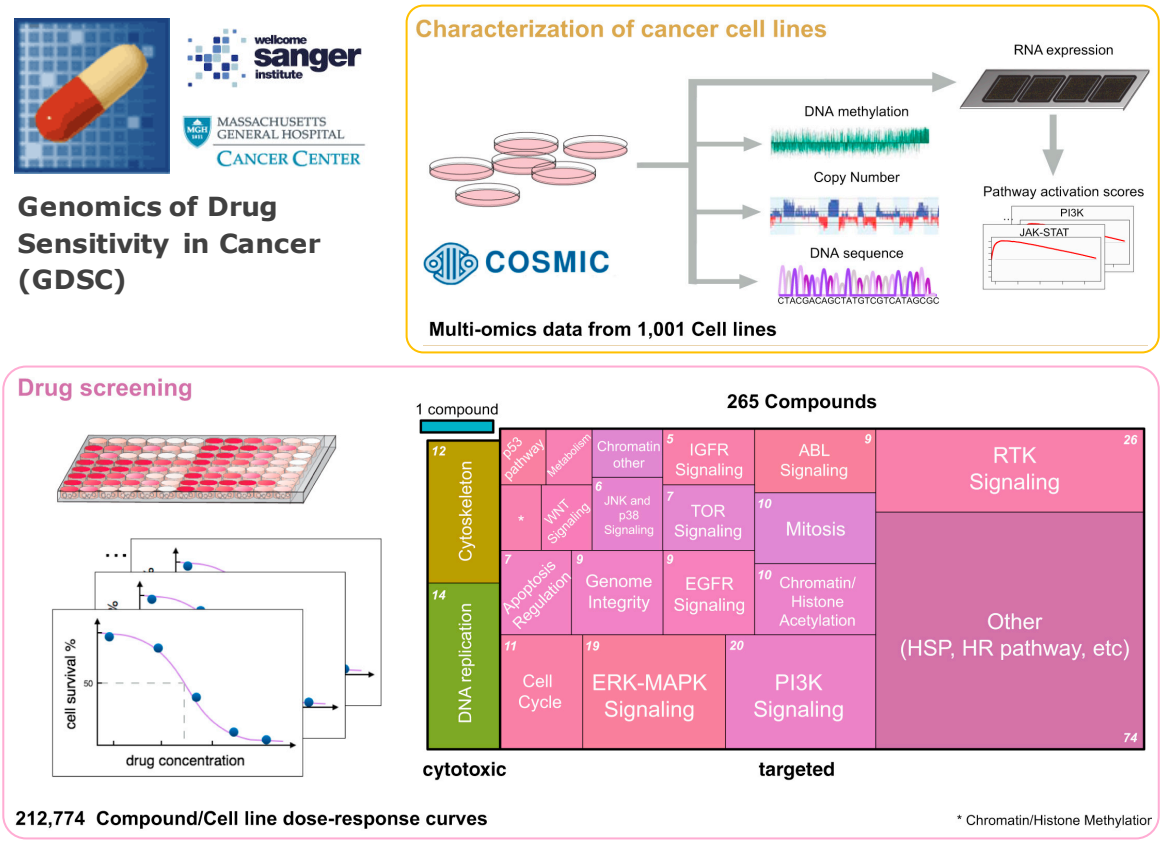


Figure 6 Genomics of Drug Sensitivity in Cancer (GDSC) dataset. Main characteristics of the GDSC dataset, a collaboration between the Wellcome Trust Sanger Institute and the Massachusetts General Hospital Cancer Center. A panel of 1001 human cancer cell lines was molecularly characterized and screened for differential sensitivity against 265 anticancer compounds.
Adapted from (Iorio et al, 2016) under a Creative Commons Attribution 4.0 International Public License.

for long-term cytotoxicity (i.e. cytostatic effects). *Membrane integrity assays* use the rupture of cellular membranes as a proxy of cell viability, and include methods such as enzyme leakage (e.g. lactate dehydrogenase, LDH) and die or fluorescent compound exclusion (e.g. propidium iodide, trypan blue). Finally, *metabolic assays*, the most commonly used, exploit cellular metabolic activity as a surrogate of cell viability (Cortes-Ciriano et al, 2016). For instance, the firefly luciferase ATP assay, used in the CCLE, CTRP and GSDC screens (**Table 1**), measures a luminiscent signal produced by the oxydation of luciferin in the presence of ATP, whose amount is proportional to the number of cells in culture (Fan & Wood, 2007; Kangas et al, 1984). In contrast, the NCI-60 screen estimated cell mass using the sulforhodamine B (SRB) assay,

a colorimetric method that relies on the property of the SRB dye to bind stoichiometrically to proteins under mild acidic conditions (Vichai & Kirtikara, 2006). Other frequently used colorimetric assays, such as the MTT or resazurin assays (also used in the GDSC), are based on the enzymatic reduction of dyes by living cells (Berridge et al, 2005).

The dose-response data obtained with these assays usually describe a sigmoidal curve that can be fit by a 4-parameter *log-logistic model* (also called Hill model) (**Figure 7a**). Hill models allow the estimation of the *half maximal inhibitory concentration* or IC_{50}, a parameter normally used to quantify compound potency that corresponds to the concentration required to reach a response midway between the minimum and maximum activities (top and bottom plateaus of the sigmoidal curve). IC_{50} values can be relative or absolute, depending on wether or not maximum responses are calculated using the actual test compound or an external positive control, such as another drug known to produce the maximum possible effect or the maximum value within a multi-compound test plate (Sebaugh, 2011). Another measure to evaluate compound activity, often preferred to IC_{50}, is the *area under the response curve* (AUC), a parameter that comprises both potency and efficacy and seems to be more robust when comparing single drugs across cell lines treated with identical dose ranges (Fallahi-Sichani et al, 2013). Other authors propose the use of alternative metrics based on *growth rate inhibition* (GR) in order to correct for confounder factors (e.g. division number) that affect conventional drug sensitivity measures (Hafner et al, 2016).

1.2.3 Computational drug sensitivity modeling

Computational techniques for the identification of molecular drug response determinants exist in a continuum of complexity. The simplest approaches correspond to purely statistical methods that aim at associating individual molecular features with drug sensitivity. For instance, *Pearson* and *Spearman's correlations* are often used in order to test the significance of the association of molecular profiles, and constitute the basis of the COMPARE algorithm developed within the NCI-60 project (Paull et al, 1989). Other standard hypothesis testing methods such as *Student's t-test*, *ANOVA* or *MANOVA* have also been applied to identify molecular features that are statistically different between sensitive and resistant cell lines (Garnett et al, 2012; Iorio et al, 2016). However, proper predictive models are needed in order

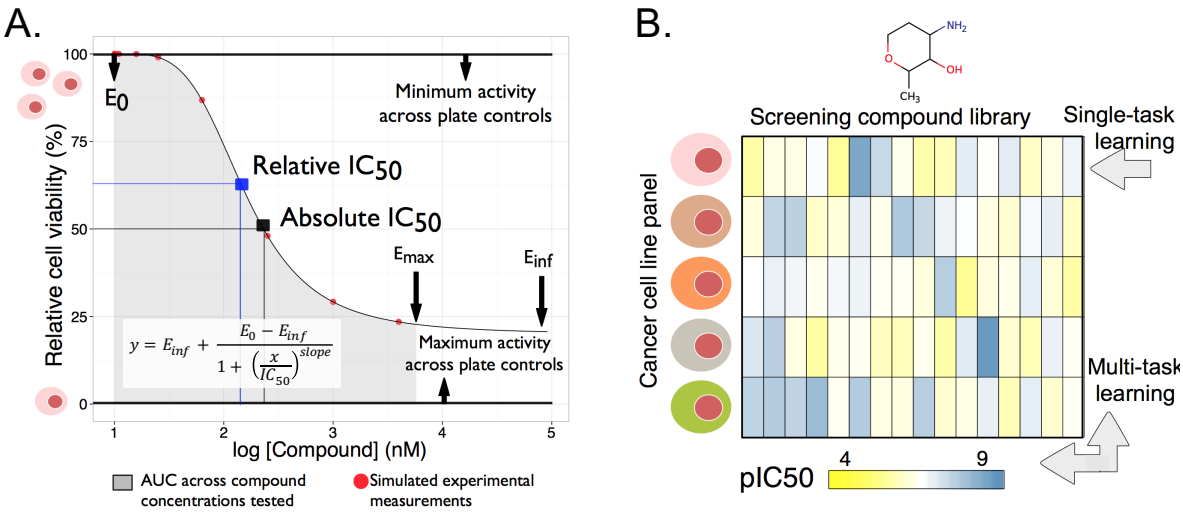


Figure 7 Dose-response data and drug sensitivity modeling. A. Dose-response data fit with a log-logistic model. The axes represents the relative cell viability in percentage (y-axis) and the compound concentration in logarithmic scale (x-axis). Red points represent simulated measurements and the black line indicates the fitted model (E_0 = 100, E_{inf} = 20, Slope = 4). The equation describing the concentration-response curve is shown in the inset. E_0 and E_{inf} are the top and bottom activity plateaus. Slope corresponds to the Hill parameter that controls the steepness of the linear part of the curve. E_{inf} indicates the effect of the compound at the maximum concentration tested. The difference between absolute and relative IC_{50} values is highlighted. AUC is indicated by the gray shade. **B.** Drug bioactivity matrix (drugs vs. cell lines). In single-task modeling, each cell line (row) or compound (column) is modeled at a time, while in multi-task learning the whole bioactivity matrix is modeled simultaneously by integrating biological and chemical information. *Adapted from (Cortes-Ciriano et al, 2016) with permission from Bentham Science Publishers LTD.*

to identify more complex relationships involving sets of interacting response markers that might act in a cumulative fashion.

These computational models of drug sensitivity typically aim at estimating a metric of drug activity (e.g. IC_{50}, AUC) for multiple single compounds on panels of cell lines using combinations of molecular features (e.g. gene expression, genomic variants, DNA methylation) of those cell lines as predictors. Such an approach is referred to as *single-task learning*, as it allows to extrapolate on the biological space (i.e. predict the response of new cell lines), but not on the chemical space (i.e. predict the effect of new compounds). In contrast, more complex *multi-task learning* approaches try to predict drug sensitivity by combining cell line and

compound information in a single model, therefore allowing extrapolation on both directions (Cortés-Ciriano et al, 2016; Gönen & Margolin, 2014; Menden et al, 2013; Zhang et al, 2015) (**Figure 7b**).

The development of predictive models can be divided in three main steps (**Figure 8**). First, data sets must be selected and preprocessed, which involves normalization and filtering of noisy or irrelevant data. Feature selection (or dimensionality reduction) might be performed at this point using different statistical methods (e.g. univariate correlation with drug sensitivity measures), but it is often part of the subsequent model training phase. In a second step, models are trained and their performance (i.e predictive power) tested. A plethora of statistical and machine learning approaches can be applied to that end, depending on factors such as the characteristics of the prediction problem or the type of input data used. Next, models are independently evaluated using datasets different to the one used for training, a process that can be iteratively repeated a number of times in order to refine the models and improve their predictive power. Finally, the best performing models are validated and, if successful, applied in the context they were created for (e.g. drug response prediction in a clinical setting). Similarly, good performing models might enable the formulation of hypotheses regarding the molecular MoA of specific compounds (Azuaje, 2017).

Input data has been reported as the dominant factor affecting model performance, with modeling methodologies playing a secondary role (Bayer et al, 2013; Jang et al, 2014). This involves both the molecular features used as predictors as well as the compound to be predicted. AUC is generally preferred to IC_{50} as a summary statistic for drug sensitivity, as it is able to capture more information about the experiment and leads to better predictive accuracy (Jang et al, 2014). SNVs, CNVs and gene expression have been extensively used as predictive features, either alone or in combination. Gene expression is by far the most widely used data type, and comparative studies have shown that it is also the most predictive (Costello et al, 2014; Geeleher et al, 2014). In contrast, while integrative models using gene expression in combination with genomic or epigenomic features might increase model performance, this increase is often only marginal (Costello et al, 2014; Jang et al, 2014). Nevertheless, genomic features possess the advantage of being more easily translatable to clinical biomarkers as well as often increasing model interpretability (Aben et al, 2016). Chemical structure, protein

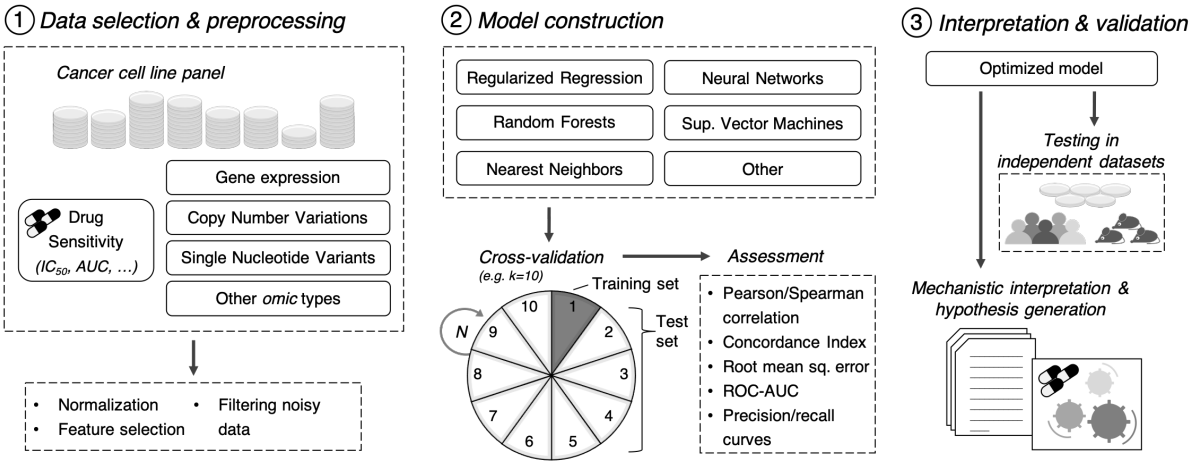


Figure 8 Development of computational models of drug sensitivity. Key steps of data selection and preprocessing, model construction and validation are shown, including the different data types, algorithms and computational methods that are most commonly used.
Based on (Azuaje, 2017).

expression, DNA methylation and pathway activity predictors are also becoming increasingly used in an attempt to improve model accuracy and interpretability (Fey et al, 2015; Menden et al, 2013; Wang et al, 2017; Zhang et al, 2015).

Most computational approaches applied to the prediction of drug sensitivity belong to the class of *supervised* learning techniques, although *unsupervised* methods such as standard clustering or principal component analysis (PCA) are extensively used in the processes of data selection and filtering (Seashore-Ludlow et al, 2015). Predictive models can also be divided into *continuous* (regression problems), which predict numerical estimates of drug sensitivity (e.g. IC_{50}, AUC) or *categorical* (classification problems), which make predictions according to predefined discrete response levels based on those descriptors (e.g. high vs. low sensitivity) (Azuaje, 2017). *Linear regression-based methods* are among the most popular to infer multivariate compound response predictors, and are discussed below in more detail. A variety of machine learning approaches have also been applied, including *random forests* (Menden et al, 2013; Riddick et al, 2011), *support vector machines* (Dong et al, 2015; Gupta et al, 2016) and *neural networks* (Menden et al, 2013), among many other. *Network-based techniques* are also being increasingly used in this and similar contexts (Qin et al, 2015; Wang et al, 2014; Zhang et al, 2015). Finally, these methods are often adapted and integrated into more

complex algorithms (Fang et al, 2015; Neto et al, 2014; Park et al, 2014), and *model ensembles* have also been used that combine predictions arising from multiple individual methods to generate a more accurate final prediction (Cortés-Ciriano et al, 2016; Wan & Pal, 2014). A number of studies have benchmarked and compared many of these approaches, generally showing a lack of general solutions to achieve good predictive power and, as mentioned above, a strong dependency on data quality and sample size (Bayer et al, 2013; Jang et al, 2014; Papillon-Cavanagh et al, 2013).

In order to assess model performance, different statistical indicators are used that measure the "goodness-of-fit" of predicted and observed values and compute its associated error. In regression, *correlations*, *coefficients of determination* and *root mean squared errors* are frequently used measures of error, while in classification problems the *areas under the receiver operating characteristic* (ROC AUC) and *precision-recall curves* are normally reported (Azuaje, 2017). Another indicator is the *concordance index* (CI), a generalization of the ROC AUC that can be applied to both continuous and categorical models (Papillon-Cavanagh et al, 2013). Moreover, predictive performance is typically assessed within a *cross-validation* (CV) framework in order to build more generalizable models and avoid overfitting (i.e. the model is so closely fit to the training data that it fails when applied to other datasets). In CV, the full dataset is divided into two subsets: a *training* set, used to build the model, and a *validation* or *test* set, used to assess its predictive performance. The most common CV scheme is the *K-fold* (KF-CV), in which the dataset is split by random sampling into K partitions (e.g. 10), of which K-1 are used for training and the remaining one for testing. The process is then iterated until all partitions have been used as test sets. Additionally, this scheme can be repeated a number of times (R times KF-CV) by creating new random splits. At the end, performance statistics of each round of validation are summarized by a single indicator (Azuaje, 2017; Baek et al, 2009).

Last but not least, interpretability is another key factor and challenge when building predictive models. Unfortunately, the most powerful models are often the least interpretable, appearing as *black boxes* in terms of the factors contributing to the final predictive decision. While achieving a good prediction accuracy is usually the main objective, it is normally a better idea to find a balance between accuracy and interpretability. In this sense, simpler methods that allow an

easy understanding of their parameters and permit the extraction of biological and mechanistic hypotheses, even at the cost of some accuracy, are often preferred (Johansson et al, 2011).

1.2.4 Regularized linear regression

Linear regression is one of the simplest and most widely used statistical methods for predictive modeling. In a standard linear regression model, the activity profile of a compound (Y) is expressed as a weighted sum of molecular features (X), where the relative contribution of each feature is controlled by a signed coefficient (β) that indicates both the strength and the direction of influence.

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \tag{1}$$

In order to fit the model (i.e. obtain the most accurate β coefficients), a *loss* function (or cost function) that calculates the prediction error needs to be minimized, being a common choice the *sum of squared errors* (SSE):

$$SSE = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \tag{2}$$

However, ordinary linear regression models do not achieve good predictive accuracy due to overfitting. To overcome this problem, so-called *regularization techniques* are used that apply a *penalty* term (P) to the loss function in order to shrink the β coefficient estimates towards zero, hence generalizing the model. The strength of the penalty is controlled by a *tuning* parameter (λ) whose value is usually optimized by cross-validation.

$$\text{regularization}: \quad SSE + \lambda P \tag{3}$$

The two most renowned regularization methods are *Ridge* (Hoerl & Kennard, 1970), which applies the sum of squared coefficients as a penalty term (also called L_2 penalty), and *LASSO* (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996), which applies the sum of absolute coefficients (L_1 penalty) (**Figure 9a**). The cost function will then become, in each case:

$$\text{LASSO}: \quad SSE + \lambda_1 L_1 = SSE + \sum_{j=1}^{p} |\beta_j| \tag{4}$$

$$\text{Ridge}: \quad SSE + \lambda_2 L_2 = SSE + \sum_{j=1}^{p} \beta_j^2 \tag{5}$$

The differences in these penalties entail a few consequences. On the one hand, LASSO is able to shrink the least important feature coefficients to zero, thus performing feature selection, while all coefficients in Ridge regression are necessarily non-zero. Therefore, LASSO yields *sparse* models that are more interpretable than Ridge models. On the other hand, LASSO does not allow *group selection* (multicollinearity): if there is a group of variables highly correlated to each other (e.g. genes participating in the same pathway or process), it tends to select one at random and ignore all the rest. Moreover, LASSO is not adequate for the $p > n$ case (i.e. more variables than samples), as it can at most select n features.

To compensate these limitations, a new regularization technique was developed as a hybrid between LASSO and Ridge, called the *Elastic Net* (Zou & Hastie, 2005) (**Figure 9b**). By linearly combining both L_1 and L_2 penalties in a proportion controlled by a *hyper-parameter* (α), Elastic Net regularization *i)* does automatic variable selection and continuous shrinkage, *ii)* is able to select groups of correlated features, and *iii)* overcomes the $p>n$ problem.

$$\begin{aligned} \text{ElasticNet}: \quad & SSE + \alpha L_1 + (1-\alpha) L_2 \quad \text{where } \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2} \\ & \text{therefore, if} \begin{cases} \alpha = 0 : \text{Ridge} \\ \alpha = 1 : \text{LASSO} \\ 0 < \alpha < 1 : \text{Elastic Net} \end{cases} \end{aligned} \tag{6}$$

Due to their high interpretability and robust performance, regularized linear regression methods, and Elastic Net in particular, have been systematically used in order to predict drug sensitivity using panels of cell lines (Barretina et al, 2012; Garnett et al, 2012; Iorio et al, 2016), and some benchmarking studies have even explicitly recommended their use for this aim (Jang et al, 2014).

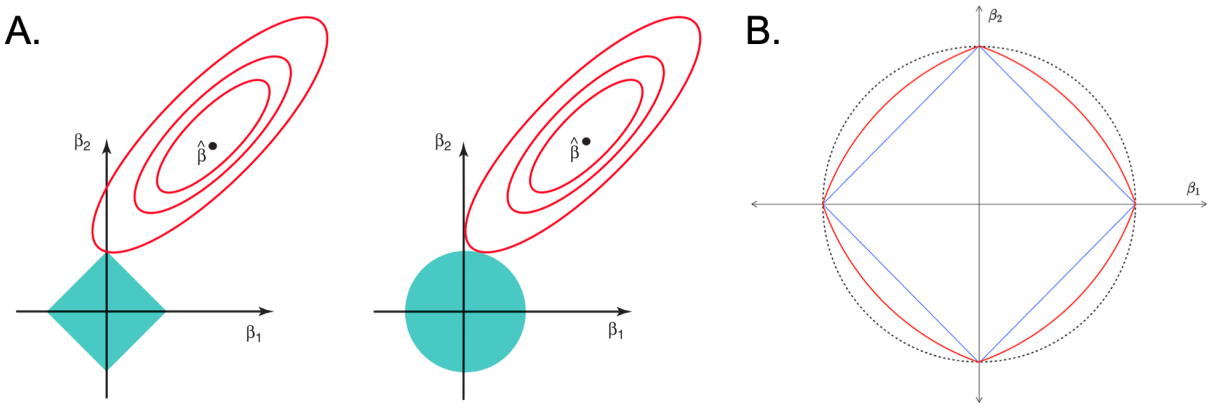


Figure 9 Geometrical representation of regularization methods. A. Contours of the error and constraint functions for the LASSO (left) and Ridge regression (right). $\hat{\beta}$ represents the least squares solution, the solid blue areas are the constraint regions of LASSO, $|\beta_1| + |\beta_2| \leq s$, and Ridge, $\beta_1^2 + \beta_2^2 \leq s$, and the red ellipses represent the regions of constant SSE. LASSO and Ridge regression coefficient estimates are given by the first point at which an ellipse contacts the constraint region. Ridge has a circular constraint with no sharp points, and therefore the intersection will not generally occur on an axis and all the coefficient estimates will be exclusively non-zero. In contrast, the LASSO constraint has corners at each of the axes, and so the intersection will often happen at an axis, shrinking one of the coefficients to zero. In higher dimensions, many coefficients may equal to zero generating a sparse model. **B.** Comparison of constraint functions of LASSO (blue), Elastic Net (red) and Ridge (black) regularization. The Elastic Net constraint corresponds to α=0.5. Tuning the value of this hyperparameter can make the Elastic Net approximate more or less to the more strict Ridge and LASSO constraints.

A was reprinted from (James et al, 2013) with permission from Springer. B was reprinted from (Korkmaz et al, 2017) with permission from Elsevier.

1.3 Functional Screening and Analysis using CRISPR-Cas9

CRISPR-Cas9 has revolutionized genome engineering, becoming in only a few years an indispensable toolbox for almost any field of biological research and presenting an invaluable potential for biotechnological and therapeutic applications (Hsu et al, 2014). As an illustration, more than 9,000 research articles have been published about CRISPR-Cas9 since its development as a genome-editing tool in 2012, and this number continues to increase every year (Adli, 2018). The unprecedented ease, precision and versatility of the CRISPR-Cas9 system compared to previous technologies such as zinc fingers or TALENs are the reasons behind its impressive success. CRISPR-Cas9 has already enabled the generation of large mutant cell line collections and the genetic modification of a variety of model organisms, and genome-wide as well as specific sgRNA libraries have been used for a diverse range of screens. Similarly, computational algorithms and tools are constantly being developed in order to facilitate the design of more efficient libraries as well as to increase the power and robustness of the analysis of CRISPR-based screens.

1.3.1 The CRISPR-Cas9 system

CRISPR-Cas modules constitute the unique adaptive immune systems of bacteria and archaea, being present in approximately 50% and 90% of their species, respectively (Grissa et al, 2007; Mojica et al, 2000). They consist in sequence specific RNA-guided endonuclease complexes that recognize and cleave foreign nucleic acids and thus protect these organisms against infection by mobile genetic elements such as phages, plasmids or transposons (Barrangou et al, 2007; Bhaya et al, 2011; Horvath & Barrangou, 2010; Marraffini & Sontheimer, 2008). The main defining characteristic of these systems, which led to their initial discovery and gives them their name, is the presence of a genomic array of *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR): a series of identical repeats separated by unique spacers of foreign origin that serve as an archive of previous infections (Bolotin et al, 2005; Ishino et al, 1987; Mojica et al, 2005; Pourcel et al, 2005). Next to this array, an operon of well-conserved *CRISPR-associated* (Cas) genes encodes proteins that are in charge of driving the immune response (Jansen et al, 2002).

The defense mechanism orchestrated by CRISPR-Cas systems can be divided in three different phases: adaptation, expression/maturation, and interference (Hille et al, 2018)(**Figure 10a**). Upon an initial infection, Cas proteins (e.g. Cas1 and Cas2) select a stretch of foreign DNA and integrate it into the CRISPR array as a new spacer. This integration happens preferentially at the end of an AT-rich *leader* sequence that precedes the whole array, therefore keeping a chronological record of infections. The CRISPR array is then transcribed into a long *precursor* CRISPR RNA (pre-crRNA) and processed within the repeats by either Cas proteins or cellular RNAses into mature crRNAs, which contain the spacer sequence and a part of the repeat. In the context of a new infection, these crRNAs guide the interference complex formed together with the Cas enzymes to their cognate foreign DNA in order to cleave it. Target recognition and cleavage requires complementary base-pairing between the crRNA spacer and the foreign sequence, and, in most cases, also the presence of a short *Protospacer Adjacent Motif* sequence (PAM), which constitutes an authentication and discrimination mechanism that avoids auto-immune targeting of the actual CRISPR array. Cleaving the DNA results in degradation of the invading virus or plasmid, thus protecting the cell against infection.

CRISPR-Cas systems are classified into two different classes, further subdivided in six types and several other subtypes (Shmakov et al, 2017). Their main difference resides in that class 1 systems (types I, III and IV), the most common in bacteria and archaea, use interference complexes made of 4-7 Cas proteins, while class 2 systems (types II, V, and VI) rely on a single multidomain protein to drive interference. This simple effector architecture has rendered class 2 CRISPR-Cas, and in particular the type II system of *Streptococcus pyogenes*, the best choice for genome engineering applications (Deltcheva et al, 2011; Garneau et al, 2010; Gasiunas et al, 2012). The effector protein in type II systems is Cas9, a large RNA-guided DNA endonuclease that contains two different nuclease domains: an *HNH domain* that nicks the strand complementary to the crRNA (target strand) and a *RuvC-like domain* responsible of cleaving the opposite strand (non-target strand). Importantly, Cas9 requires for its activity the crRNA together with a second non-coding RNA called the *trans-activating* RNA (tracrRNA), which is partially complementary to the crRNA and warrants the stability of the interference complex. Through its endonuclease activity, Cas9 generates a blunt DNA *double-strand break* (DSB) 3 bp upstream of a NGG specific PAM sequence (Jiang & Doudna, 2017).

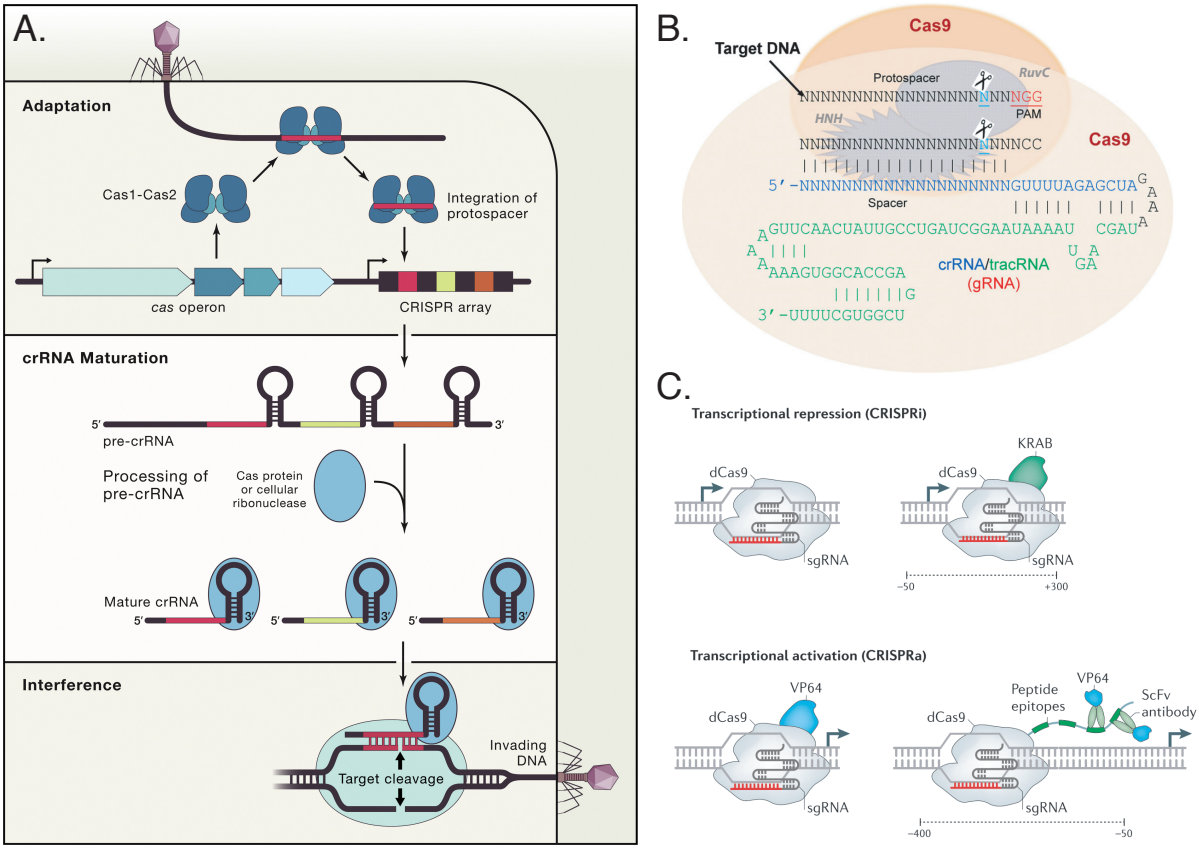


Figure 10 The CRISPR-Cas9 system. A. Schematic representation of the three phases during CRISPR-mediated immunity. **B.** Structure of the complex formed between Cas9 and the chimeric guide RNA (gRNA). crRNA/tracrRNA and PAM sequences are indicated in different colors. Sequences corresponding to the spacer and protospacer are represented by stretches of N nucleotides. HNH and RuvC domains are depicted in grey. **C.** dCas9-mediated transcriptional repression and activation. Alternative strategies for both CRISPRi and CRISPRa are summarized. CRISPRi can be achieved by dCas9 alone through steric hindrance or by a fusion with a transcriptional repressor such as KRAB. Two options for CRISPRa are the fusion of dCas9 to VP64 or, alternatively, to an array of repeating peptide epitopes that recruit multiple copies of single-chain variable fragment (ScFv) antibodies fused to transcriptional activation domains.

A was reprinted from (Hille et al, 2018) with permission from Elsevier, B and C were adapted from (Khalili et al, 2015) and (Shalem et al, 2015), respectively, with permission from Springer Nature.

The breakthrough came with the demonstration that the CRISPR-Cas9 system can be easily exploited for genome editing in eukaryotic cell lines (Cong et al, 2013; Jinek et al, 2012; Mali et al, 2013). By including 20 nucleotide long custom spacers into the crRNA sequence, any gene or genomic locus can be specifically targeted provided it contains a 5'-NGG-3' PAM downstream

of the protospacer sequence. To simplify the system, crRNA and tracrRNA were fused into a single chimeric guide RNA (gRNA, or sgRNA) that proved to be as efficient in generating DSBs as the two separate RNAs (Jinek et al, 2012) (**Figure 10b**). The repair of these DSBs by eukaryotic cellular mechanisms generally involves the *non-homologous end joining* (NHEJ) pathway, which is error-prone and leads to the introduction of small insertions and deletions (indels). Indels in coding regions that cause frameshits or premature stop codons result in gene inactivation and the generation of a stable *knock-out* (KO) (Brinkman et al, 2018).

In the last years, the CRISPR-Cas9 system has been tuned for a plethora of applications. For instance, a catalytically inactive form of Cas9 (dCas9) has been used as a recruitment platform for other effector proteins and RNA aptamers in order to achieve targeted transcriptional activation (CRISPRa) (Konermann et al, 2015), transcriptional inhibition (CRISPRi) (Qi et al, 2013), epigenome editing (Hilton et al, 2015; McDonald et al, 2016) and base editing (Komor et al, 2016), among other (**Figure 10c**). Also, the use of donor DNA templates in modified Cas9 systems allows the introduction of specific mutations or *knock-ins* in particular loci via *homology-directed repair* (HDR) (Aird et al, 2018). Moreover, alternative Cas9-like proteins from species different to *S.pyogenes* that differ in size, PAM requirement or cleavage mechanism have been explored for different uses. Finally, many efforts are being dedicated into re-engineeing the already well-characterized Cas9 proteins in order to reduce size, increase fidelity, reduce off-target effects, and extend targeting scope (Adli, 2018). Further improvements will therefore continue to expand the uses and applications of the CRISPR-Cas9 system.

1.3.2 CRISPR-Cas9-based pooled genetic screens

Pooled genetic screens are cost-effective means for interrogating genomes in search of causative links between genotype and phenotype. Due to its simplicity, scalability, efficiency and low off-target effects compared to previous methods (e.g. RNA interference), CRISPR-Cas9 has become an incredibly practical tool for genetic screening, and a number of genome-wide knock-out, inhibition and activation sgRNA libraries have already been generated and successfully tested (Sanson et al, 2018).

In a pooled screen, thousands of individual perturbations (sgRNAs, in this case) are simultaneously tested on a population of cells, which are then selected for a robust phenotype (e.g. survival) and the relevant perturbations identified by comparison with unselected samples. Two main modalities of pooled screens can be distinguished. *Negative* selection or *dropout* screens aim at identifying perturbations that cause the depletion of cells over time, and have been widely used for the characterization of essential genes (essentialomes) and tumor cell dependencies (Hart et al, 2015; Wang et al, 2015; Wang et al, 2017). Such screens face the challenge of distinguishing dropout due to biological reasons from random depletion, and hence require the use of significantly large cell numbers in order to guarantee representation as well as robust statistical analyses (Doench, 2018; Shalem et al, 2015). In contrast, *positive* screens expose cells to a strong selective pressure (e.g. a compound, an infection, a different culture condition) that allows the survival of only those that carry resistance-conferring perturbations. Given that the majority of the population is eliminated, resistant cells are strongly enriched, and therefore positive screens present a much larger dynamic range and are generally more robust than negative screens (Doench, 2018; Shalem et al, 2015). This type of screen has been effective for the identification of host factors required for viral infection (Marceau et al, 2016) and genes involved in drug response (Deans et al, 2016; Koike-Yusa et al, 2014; Zhou et al, 2014).

The first step in CRISPR-Cas9 knock-out genetic screening consists in the design of the custom sgRNA library that will be used to perturb the cell population (**Figure 11**). In these libraries, every gene is typically targeted by several sgRNAs in order to ensure coverage and achieve higher statistical certainty when analyzing screen results, and 6 sgRNAs per gene has been reported as good compromise between performance and library size (Ong et al, 2017). sgRNA specificity and activity is highly influenced by the actual nucleotide sequence as well as the surrounding target site, and although this is still an active area of study, a few features have been described that can be taken into consideration when designing sgRNAs (**Figure 11b**). As previously explained, the sgRNA target sequence must be located 5' to a PAM site, and purine nucleotides seem to be preferred at the position immediately adjacent to it (position 20) (Gagnon et al, 2014). Similarly, sgRNAs are normally designed to target 5' exons, as this increases the chances of generating deleterious frameshifts, although targeting conserved protein domains

has also proven successful (Shi et al, 2015). In contrast, sites upstream of alternative start codons, containing high-frequency SNPs or subject to alternative splicing should be avoided (Doench et al, 2014). Moreover, while the sgRNA spacer sequence needs to be strictly identical to the ~10 bases of the target closest to the PAM (the *seed* sequence), mismatches are allowed at the 5' end, which should be considered to evaluate possible off-target effects in regions of high similarity. Other characteristics such as GC content, DNA strand or local chromatin structure have also been reported to affect sgRNA activity (Doench et al, 2014; Wang et al, 2014), and sgRNA scaffolds have been subject to optimization in order to improve efficiency (Dang et al, 2015). These and other considerations have been implemented into different prediction algorithms and CRISPR tools that facilitate sgRNA design with the objective of finding the optimal trade-off between on-target efficiency and off-target effects (Doench, 2018; Haeussler et al, 2016; Heigwer et al, 2014).

Once designed *in-silico*, sgRNAs are synthesized as a pool of oligonucleotides using array-based methods (Kosuri & Church, 2014), amplified, and cloned to generate the plasmid library that will be used for screening. Lentiviruses or other retroviruses are generally used for Cas9/sgRNA delivery due to their ability to integrate into the genomes of target cells, marking them permanently with the perturbation and allowing subsequent readout by *next-generation sequencing* (NGS). Single and dual-vector systems have been developed that are codon optimized and include all necessary elements (e.g. promoters, UTRs) to ensure proper expression in the target eukaryotic cells (**Figure 11c**). An antibiotic resistance marker (e.g. puromycin) is usually also included in order to allow the selection of successfully transduced cells (Sanjana et al, 2014). While delivering Cas9 and sgRNAs together within the same vector is generally the preferred option, it can also be challenging as viral titers tend to be low due to the large size of the Cas9 gene (Shalem et al, 2015).

Transduction of the target cells with the lentiviral library needs to be performed in conditions that maximize infection, therefore requiring fewer cells to initiate the screen, while ensuring that the majority of them receives only a single sgRNA, as multiple sgRNAs per cell would lead to confounding and unspecific effects. Such conditions are usually obtained with a *multiplicity of infection* (MOI) of 20-40%, as illustrated by the corresponding Poisson distribution (Doench, 2018) (**Figure 11d**). Furthermore, the number of cells utilized for the screen needs to allow a

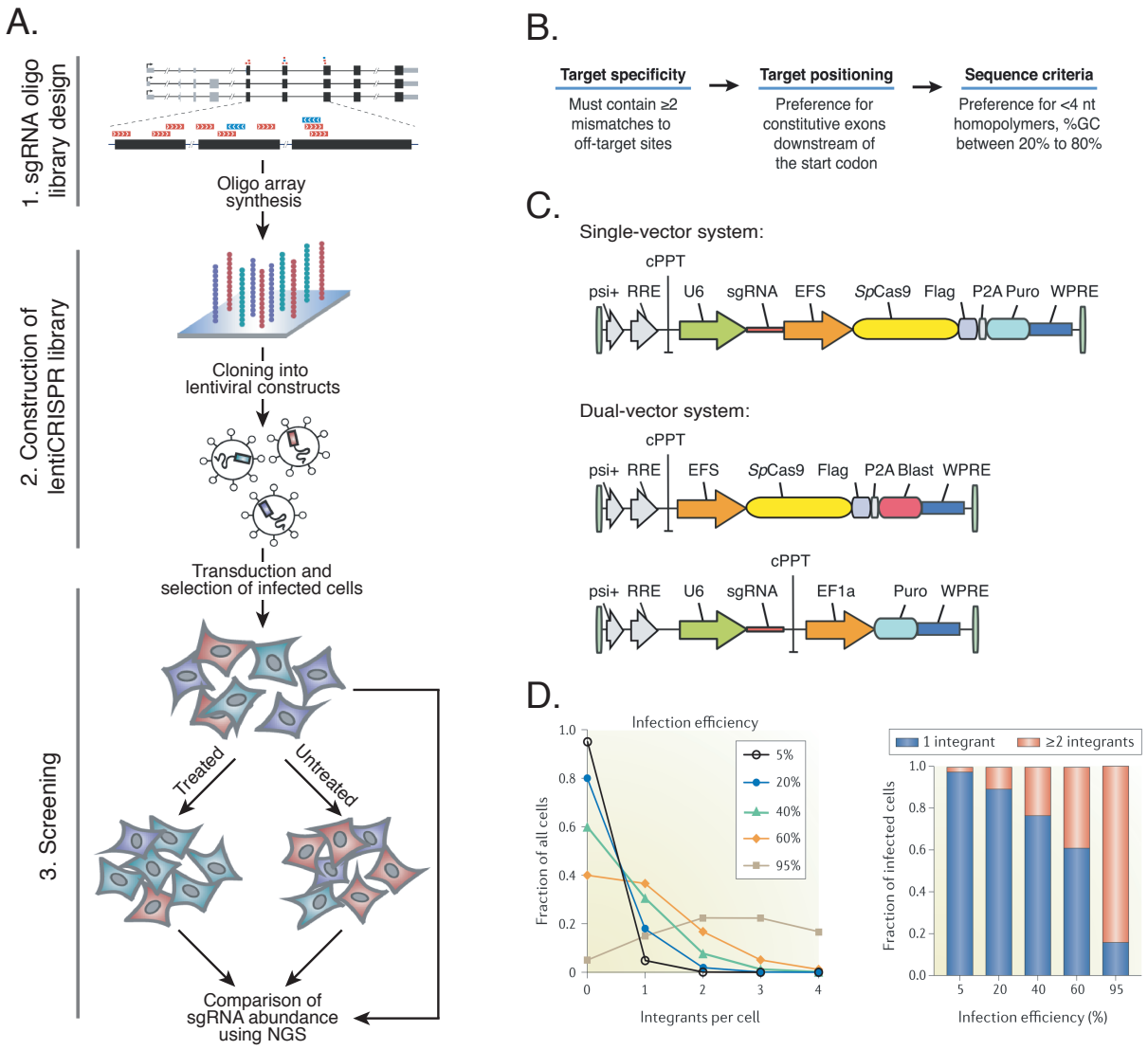


Figure 11 CRISPR-KO library design and genetic screening. A. General pipeline of a pooled positive genetic screen using a lentiviral CRISPR-Cas9 system.**B.** Example of sgRNA design, including some criteria used for sgRNA selection. **C.** Single (lentiCRISPRv2) and dual (lentiCas9-Blast, lentiGuide-Puro) lentiviral systems for the delivery and expression of *S. pyogenes* Cas9 (SpCas9) and sgRNA components. psi+: Psi packaging signal; RRE: Rev response element; cPPT: central polypurine tract; U6: U6 promoter for RNApol III; EFS: elongation factor 1a short promoter; EF1a: elongation factor 1a promoter; Flag: Flag octapeptide tag; P2A: 2A self-cleaving peptide; Puro: puromycin selection marker; Blast: blasticidin selection marker; WPRE: post-transcriptional regulatory element. **D.** Poisson distribution for the fraction of cells carrying varying numbers of sgRNAs at different infection efficiencies (left) and barchart of the percent of infected cells with 1 or more sgRNAs at different infection efficiencies (right).

A and B were adapted from (Wang et al, 2014) with permission from The American Association for the Advancement of Science, and C and D were adapted from (Sanjana et al, 2014) and (Doench, 2018), respectively, with permission from Springer Nature.

high coverage of the library i.e. every sgRNA being represented in the target cells a sufficiently high number of times (e.g. 1000x). After selection of infected cells and exposure to the stimulus of interest for a few days (e.g. a drug, in positive screens), genomic DNA is isolated from cells, amplified, and sgRNAs are sequenced using NGS. Specific adapters and barcodes are added during PCR amplification in order to allow multiplexing strategies during sequencing, such as double indexing (Kircher et al, 2012). Finally, comparison of sgRNA representation between treated and untreated samples (positive screens) or between samples at different time points (negative screens) using dedicated statistical methods allows the obtention of "hit lists" containing candidate genes that can be further validated.

1.3.3 Computational analysis of CRISPR-Cas9 knock-out screens

The statistical analysis of CRISPR-Cas9 screens starts with the generation of a count table of sgRNA abundance across samples. Adapted *RNA-seq* pipelines as well as dedicated CRISPR tools can be used for preprocessing raw NGS reads (i.e. fastq files), which involves the steps of sample demultiplexing, trimming of library-specific adapters, and mapping guide sequences to the original sgRNA library. One or two mismatches or small shifts in sgRNA sequences are sometimes allowed during mapping in order to account for sequencing errors and increase yield, although strict matching is generally preferred to avoid false positives (Dai et al, 2014). Sample quality can be evaluated by analyzing the number and percentage of mapped reads and the distribution of sgRNA counts. A sufficient number of counts for each sgRNA across samples (e.g. over 300) together with a low value of zero-count sgRNAs in the initial library and time points are required in order to achieve good statistical power during subsequent analysis. Low percentages of mapped reads can be indicative of NGS and oligonucleotide synthesis errors as well as sample contamination. Similarly, unbalanced oligonucleotide synthesis, low transfection efficiencies and overselection can lead to highly dispersed count distributions in plasmid libraries, initial time points or negative selection screens, respectively. Only in positive selection screens, due to the dominance of a small number of clones over the whole population, it is common to observe higher variances in count distributions. Pairwise correlations and other exploratory analyses such as *Principal Component Analysis* (PCA) are also commonly performed in order to assess replicate consistency (Li et al, 2015).

A few challenges need to be overcome when analyzing sgRNA count data. The first of them consists in the low number of replicates present in most screens together with the highly variable and over-dispersed nature of read count distributions, which requires the use of robust normalization methods and proper statistical models in order to estimate variances and calculate the statistical significance of sample comparisons (Anders & Huber, 2010). As this is a common problem to other NGS experiments, existing algorithms for RNA-seq differential expression analysis, sometimes with small adaptations, are normally used (Dai et al, 2014; Hardcastle & Kelly, 2010; Love et al, 2014). These methods generally calculate a normalization factor to correct for library size, and most of them use the *Negative Binomial* (NB) distribution to model read counts, as this distribution is able to accommodate over-dispersion and has been shown to achieve better specificity and sensitivity than other models (Di et al, 2011; Rapaport et al, 2013). Then, specific significance tests are used to compare treatment and control samples or different time points, providing a final list of sgRNAs ranked by their associated *p-values*. A more detailed explanation for the DESeq2 method is given below (Love et al, 2014).

A second challenge is related to the actual nature of CRISPR libraries. Given that several sgRNAs of different specificities and efficiencies are used to target every gene in the library, an aggregation method is needed in order to obtain a final ranked list of candidate genes. Traditional methods that combine p-values (e.g. *Fisher's method*) or z-scores (e.g. *Stouffer's method*) can be used to this end (Won et al, 2009; Zaykin, 2011). However, the most widely used approaches are based on testing the enrichment of sgRNAs that target the same gene at the top of the ranked list, which is the same principle behind traditional gene set enrichment tools (Subramanian et al, 2005; Wu et al, 2010; Wu & Smyth, 2012). Some of these methods were actually developed for RNAi screens, as they also present the same problem (Birmingham et al, 2009). This is the case of *RNAi Gene Enrichment Ranking* (RIGER), a method based on *Gene Set Enrichment Analysis* (GSEA, explained below) (Luo et al, 2008), and *Redundant siRNA Activity* (RSA), which uses an iterative hypergeometric test (König et al, 2007). Other more sophisticated techniques, such as *HitSelect*, have been specifically developed for CRISPR screens (Diaz et al, 2015).

Another often ignored source of false positives in CRISPR-based knock-out screens is the effect of targeting regions affected by *Copy Number Variations* (CNV). In these cases, sgRNAs

direct Cas9 to every copy of the targeted locus, inducing multiple DSBs and extensive DNA damage, which leads to cell depletion in a gene function-independent manner (Aguirre et al, 2016; Sheel & Xue, 2016). A few methods have been developed to correct this bias, and they can be used either alone (Meyers et al, 2017) or integrated into screen analysis pipelines (Wu et al, 2018).

CRISPR-specific tools have been developed in the last years. *MAGeCK* (Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout) is probably the most well-known, and currently exists in two different flavors (Li et al, 2015; Zhang et al, 2018). MAGECK-RRA follows a very similar approach to DESeq2 for count modeling and hypothesis testing, and then applies a modified *Robust Ranking Aggregation* algorithm (α-RRA) to aggregate guides and rank genes (Li et al, 2014). In contrast, MAGECK-MLE uses maximum likelihood to estimate gene essentiality with a probabilistic mixture model that includes guide efficiencies, therefore eliminating the need of a subsequent sgRNA aggregation step, and is able to account for more complex experimental designs (Li et al, 2015). Other existing methods, such as *BAGEL* (Bayesian Analysis of Gene EssentiaLity) (Hart & Moffat, 2016) and *ScreenBEAM* (Yu et al, 2016), use Bayesian statistics to fit models that also consider all sgRNAs targeting the same gene at once.

Finally, dedicated python and R libraries, such as *PinAPL-Py* (Spahn et al, 2017) and *Carpools* (Winter et al, 2016), and integrative web applications, such as *CRISPR analyzer* (Winter et al, 2017), offer the possibility to perform complete workflows, from NGS data processing and quality assessment to statistical screen analysis, evaluation and visualization, and provide a large range of method choices (including the ones explained above) that can be compared.

1.3.4 DESeq2 and GSEA for the analysis of CRISPR screens

DESeq2 (Love et al, 2014) is a good example of a robust method that has already been applied to the analysis of CRISPR-Cas9 KO screens (Sedlyarov et al, 2018; Seyednasrollah et al, 2015; Winter et al, 2017). This algorithm first normalizes samples to account for library size using the *median-of-ratios* method (Anders & Huber, 2010) and, as mentioned above, models

counts using the NB distribution. As the estimation of within-group dispersions (i.e. among replicates) is typically hindered by low replicate numbers, DESeq2 takes the approach of pooling features (sgRNAs, in this case) of similar mean value together for its calculation, assuming a direct mean-variance relationship. To this end, it uses *maximum likelihood* (ML) to get an initial estimation of dispersion for each sgRNA, then fits a smooth curve for the estimations dependent on the average counts, and finally shrinks sgRNA-specific dispersions towards the fitted values using a *Bayesian* approach. Such shrinkage procedure not only decreases high variances but also raises underestimated dispersions, thereby reducing the chances of false positives. Only unusually extreme gene-wise estimates (i.e. more than 2 standard deviations above the curve) are left unshrunken, as these could obey to biological reasons.

Next, *generalized linear models* (GLM) are fitted for each gene, which provide the flexibility to treat both simple and complex designs, such as the inclusion of different time points and treatments. A common problem of dealing with count data is the tendency to overestimate *log-fold changes* (LFC) for features with low counts, which can then lead to false positives (Bottomly et al, 2011). To overcome this issue, DESeq2 applies a second empirical Bayesian procedure that shrinks LFC estimates towards zero and whose effect is stronger for features with lower information (e.g. lower counts). Finally, a *Wald test* is used to calculate the statistical significance of LFCs being different than zero, and p-values are corrected for multiple testing using *Benjamini-Hochberg's* method (Benjamini & Hochberg, 1995). In order to sort sgRNAs in the final list, shrunken LFCs seem to be a reasonable choice, as these estimates are more robust than standard LFCs and represent biological effects more closely than p-values.

Gene Set Enrichment Analysis (GSEA) can be then be used to to aggregate sgRNAs into genes (Sergushichev, 2016; Subramanian et al, 2005). The aim of this algorithm is to determine if members of a gene set (in this case, sgRNAs targeting the same gene) are randomly distributed through the the ranked list or if they significantly accumulate at its beginning. To this end, first an *Enrichment Score* (ES) is calculated for each gene. Such score is the maximum deviation from zero of a *running-sum* statistic that walks down the ranked list and increases when a sgRNA targeting that gene is found (with a magnitude depending on the LFC) and decreases otherwise. Then, a *permutation* test is performed in order to estimate the statistical significance of the ES: sgRNA labels are randomly reassigned and ES recomputed in an

iterative manner to generate a null distribution, and the p-value of the real ES is calculated relative to it. Finally, p-values are adjusted for multiple testing.

A drawback of this method is that genes for which sgRNAs concentrate in the middle or at the bottom of the list can also get significant ESs. In order to overcome this problem, the initial list of sgRNAs obtained from DESeq2 needs to be first filtered by applying a threshold on p-values (e.g. p-value$\leq$0.05) so that only significant guides, ranked by LFC, will be passed on to the GSEA algorithm. A second filter on the minimum number of significant guides that is required to assess gene enrichment can also be recommended (e.g. at least two sgRNAs). The final result is a list of candidate genes, ranked by *adjusted p-values* (False Discovery Rate, FDR) and annotated with ES and total number of guides contributing to the enrichment, which can be used for further validation.

1.4 Aims of this thesis

The main purpose of this thesis was to systematically explore the large and heterogenous family of *Solute carriers* (SLCs) in relation to their role as mediators of drug disposition and modulators of drug action. In particular, we aimed at:

1. Assessing and quantifying the understudied nature of SLCs in comparison to other gene groups, considering its causes and proposing experimental solutions.
2. Contributing to the deorphanization and functional characterization of SLCs, at least in regard to drug action.
3. Investigating the mechanism of transport and transporter dependencies of a set of diverse chemical compounds of high therapeutic interest.
4. Uncovering new SLC dependencies of drug sensitivity by a combination of experimental and computational techniques, which involved:
    a. the application of computational methods for drug sensitivity prediction in cancer cell lines based on transporter molecular data.
    b. the development of a CRISPR-based genetic screening approach and its corresponding analysis pipeline for SLC genes.

Results

2

2.1 Prologue

Solute carriers (SLCs) constitute the largest group of membrane transporters in the human genome and play an essential role in cellular metabolism as mediators of nutrient uptake, waste removal and pH, ion and volume homeostasis (Hediger et al, 2013). In addition, they are highly relevant for pharmacology due to their druggability and known implication in drug disposition, and many of them are associated with disease (DeGorter et al, 2012; Rask-Andersen et al, 2014). Yet, in spite of their importance, the majority of SLCs is functionally uncharacterized and almost no systematic, integrative studies have been reported to date.

In the article presented below, we used publicly available information to demonstrate the understudied nature of SLCs as a group and to support the necessity of integrative multi-omic studies to characterize them. We first analyzed the literature in order to quantify "SLC knowledge", revealing that SLCs are the group of human genes with the largest publication asymmetry, with a few members garnering thousands of publications while most of them being barely studied, in contrast to other membrane protein families such as GPCRs and ABCs. Similarly, database interrogation showed that more than three fourths of SLCs with an already identified disease link have no active compounds associated with them. We then reviewed the historical and technical reasons behind this lack of research attention and proposed experimental approaches that could be used to de-orphanize the "SLCome". Finally, we analyzed SLC expression patterns across tissues and cell lines and derived an SLC co-expression network that illustrates the coordinated function and interdependencies of these proteins and shows their regulation in robust tissue-dependent modules.

2.2 Manuscript #1

César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM & Superti-Furga G (2015) A Call for Systematic Research on Solute Carriers. Cell 162(3): 478-87 doi: 10.1016/j.cell.2015.07.022

Detailed author contribution: A. César-Razquin performed all the analyses, designed all figures and contributed to writing the manuscript; B.S. helped designing analyses and figures and contributed to writing the manuscript; T. F-B, R.I. and X.B. contributed to literature analysis; G.G. provided support for bioinformatic analysis; D.H. assembled and curated the list of compounds targeting SLCs; M.A.H. and A.M.E. co-supervised all the work; G.S-F. supervised the work and wrote the manuscript. All co-authors read and commented on the manuscript.

A Call for Systematic Research on Solute Carriers

Adrián César-Razquin,[1,8] Berend Snijder,[1,8] Tristan Frappier-Brinton,[2] Ruth Isserlin,[3] Gergely Gyimesi,[4] Xiaoyun Bai,[5] Reinhart A. Reithmeier,[5] David Hepworth,[6] Matthias A. Hediger,[4,*] Aled M. Edwards,[2,*] and Giulio Superti-Furga[1,7,*]

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria
[2]Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada
[3]The Donnelly Centre, University of Toronto, Toronto, Ontario, M5S 3E1, Canada
[4]Institute of Biochemistry and Molecular Medicine and Swiss National Center of Competence in Research, NCCR TransCure, University of Bern, 3012 Bern, Switzerland
[5]Department of Biochemistry, University of Toronto, Toronto, Ontario, M5S 1A8 Canada
[6]Worldwide Medicinal Chemistry, Pfizer Worldwide Research and Development, Cambridge, MA 02139, USA
[7]Center for Physiology and Pharmacology, Medical University of Vienna, 1090 Vienna, Austria
[8]Co-first author
*Correspondence: gsuperti@cemm.oeaw.ac.at (G.S.-F.), matthias.hediger@ibmm.unibe.ch (M.A.H.), aled.edwards@utoronto.ca (A.M.E.)
http://dx.doi.org/10.1016/j.cell.2015.07.022

Solute carrier (SLC) membrane transport proteins control essential physiological functions, including nutrient uptake, ion transport, and waste removal. SLCs interact with several important drugs, and a quarter of the more than 400 SLC genes are associated with human diseases. Yet, compared to other gene families of similar stature, SLCs are relatively understudied. The time is right for a systematic attack on SLC structure, specificity, and function, taking into account kinship and expression, as well as the dependencies that arise from the common metabolic space.

Individual cells, be they prokaryotic or eukaryotic, must control chemical exchange with their environments, and they use lipid membranes and proteinaceous channels and transporters to this end. The lipid environment of the membrane prevents intrusion or leakage into the *sancta sanctorum* of the inner milieu and buffers the cell against changing and noxious environmental conditions, as well as against attack by phages, viruses, or bacteria (Köberlin et al., 2015; Mulkidjanian et al., 2009). In many respects, the integrity of the membranes represents as critical an element to cellular individuality as does the preservation and transmission of genetic information (Schrum et al., 2010). The protein components of cell membranes import and export most of the chemical matter essential for life, including water, ions, gases, nutrients, vitamins, cofactors, and many drugs (Kell et al., 2011; Kell and Oliver, 2014; Lin et al., 2015). Therefore, regulation of small-molecule transport across membranes is key to a cell's internal physiology and is the gatekeeper to its interface with the environment (Nigam, 2015). Yet, despite their central role in mediating the discussion between chemistry and biology and despite the fact that ~10% of the human genome encodes for transport-related functions (Hediger et al., 2013), transporters, as a class of proteins, do not appear to garner quite the attention that they deserve.

Transporters comprise solute carriers, ion channels, water channels, and ATP-driven pumps, including ABC transporters. Of these, the largest group is formed by the solute carrier proteins (SLCs), which according to the current counting comprises 456 members, distributed in 52 subfamilies that can be further phylogenetically grouped (Hediger et al., 2013, 2004; Schlessinger et al., 2010, 2013b). SLCs are membrane integral proteins localized on the cell surface and in organellar membranes and comprise facilitative transporters, which are equilibrative, and secondary active transporters (symporters and antiporters), which may be concentrative (Hediger et al., 2013). After G-protein-coupled receptors (GPCRs), SLCs are the second-largest family of membrane proteins in the human genome (Hoglund et al., 2011). For detailed information about the individual SLC family members, please refer to www.bioparadigms.org.

Links to Therapeutics and Human Disease

Much research on SLCs has been spurred by their relevance to pharmacology and drug discovery, either as drug targets themselves or as mediators of drug disposition. Drug targets include SLC6A4 (SERT), the target of the hugely important serotonin uptake inhibitor drug class. Mediators of drug transport include SLCO1B1, which transports statins and allows for preferential drug distribution into the liver compared to other tissues, such as muscle. This tissue distribution of statins is important in driving their therapeutic index by increasing the lipid lowering over the myopathy-causing activity (Giacomini et al., 2010).

SLC-mediated transport of statins and other drug classes can also render their pharmacokinetics susceptible to drug-drug interactions. For example, naringin from citrus fruits inhibits the enterohepatic transporter SLCO1A2 and thus can reduce the bioavailability of drugs that rely on this transporter, such as fexofenadine (Bailey, 2010). Transport can also be affected by the natural pharmacogenomic variability in SLCs (Giacomini et al., 2013). Other SLCs have been studied for their roles in physiology, like SLC25A7 (UCP1), the mitochondrial uncoupling protein involved in the thermogenesis process of brown adipose tissue.

Newer research has implicated SLCs in the action of chemotherapeutics; YM155, a cancer drug in clinical evaluation, was found to be completely dependent on SLC35F2 for entry into human tumor cells (Winter et al., 2014). Increasingly, SLCs are attracting attention because they mediate drug-drug and nutrient-drug interactions. For instance, the investigational

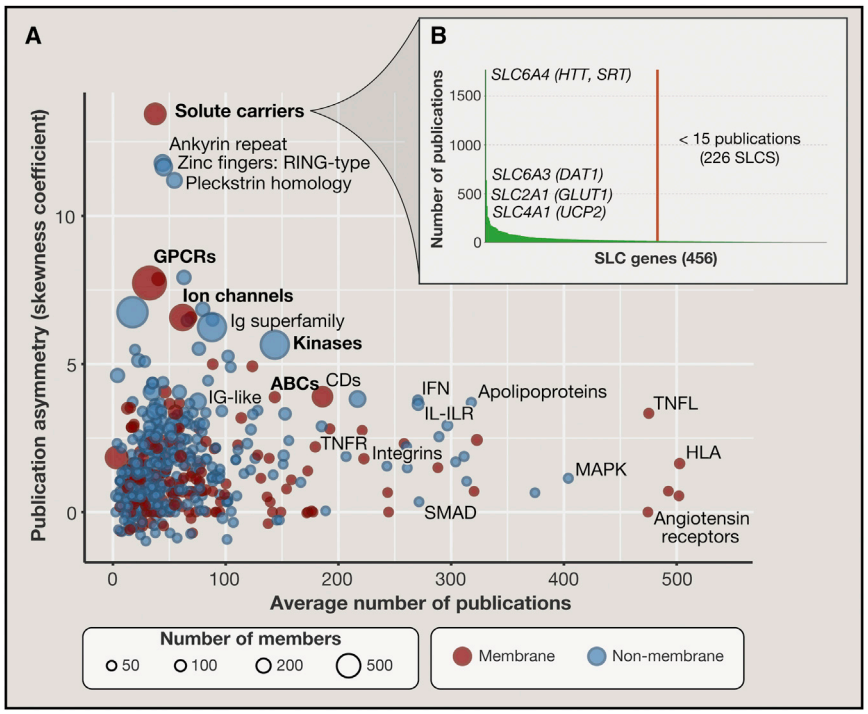


Figure 1. SLCs Are the Most Neglected Group of Genes in the Human Genome

(A) Publication asymmetry is plotted against the average number of publications per group of genes. Publication counts per gene were retrieved from the gene2pubmed file provided and curated by NCBI. Gene groups comprise all HGNC gene families and super-families as well as the GO annotations for kinase activity ("kinases") and ion channel activity ("ion channels"). Asymmetry is measured for each group of genes by calculating the skewness (as implemented in R's "moments" package) of the distribution of the number of publications for all genes within the group. A very positive skew thus indicates an uneven distribution where a few genes in the family concentrate a much higher number of publications than the rest. Dot size relates to the number of members in each gene group, and color indicates gene groups where at least 80% of their members are annotated as membrane proteins by GO annotation (see legend). Labels for selected classes are shown.

(B) Number of publications per SLC gene is displayed in descending order. The four SLCs with the most publications are annotated. The red line indicates the border at which genes have fewer than 15 publications.

JAK2 inhibitor fedratinib, which was recently terminated from development due to incidence of Wernicke's encephalopathy during trials, has been shown to inhibit thiamine uptake mediated by SLC19A2 (hTHTR2), possibly contributing to the offside effects (Zhang et al., 2014). It would not be surprising if further unplanned SLC-drug interactions were uncovered in the future.

There is also growing interest in SLCs because of their clear genetic link to human diseases; about 190 different SLCs have been found mutated in human disease and through genome-wide association studies (Williams et al., 2012, 2014).

Are SLCs Getting the Attention They Deserve?

Our sense was that the SLC protein family, despite its clear relevance to health and disease, was comparatively less well studied than other gene families. In an attempt to quantify "SLC knowledge" versus other gene families, we surveyed the literature and analyzed the distribution of publications as reported by NCBI for each gene family annotated by HGNC in an automated, unbiased fashion (Bruford et al., 2008). We then visualized the publication asymmetry, defined by the coefficient of skewness, versus the average number of publications for each family (Figure 1A). SLCs show by far the greatest publication asymmetry of all gene families, i.e., the most uneven distribution of papers over the group members. This does not seem to be simply due to a bias against membrane proteins in general, as ABC proteins, ion channels, and GPCRs appear not so unevenly distributed. Further, SLCs have an average number of publications per member of around 35, which is half of what is observed on average over all families (66 publications). At the other end of the spectrum, one finds, among others, that the small TNF superfamily of ligands are all equally and very well studied.

We then analyzed the asymmetry within the SLC knowledge domain. We performed an automated search for publications per each of the 456 SLC genes (including 65 pseudogenes), which indeed displayed a highly skewed SLC knowledge distribution curve (Figure 1B). A manually annotated search revealed the same general pattern (Figure S1B). Both analyses reveal that some gene members are extremely well studied, whereas most have very few publications. In a phenomenon that appears to be general to all human protein families, the most well-studied SLCs in the last 2 years are almost identical to those that were the most well studied a decade ago (Edwards et al., 2011). Prior to 2003, 20 of the ~400 SLC family members accrued 29% of the publications for the entire family, and those exact same family members garnered 32% of all SLC publications over the period 2012–2014 (Figure S1A).

Rankings of the SLC family members do not seem to be indicative of biological relevance. Some of the most well-studied SLCs appear to have become objects of investigation simply due to their abundance and tissue-specific expression in easily isolated cell types, which greatly facilitated their study in the era before molecular biology. Examples of this type include the so-called "band 3 of erythrocytes" protein (SLC4A1) and the erythrocyte glucose transporter GLUT1 (SLC2A1).

An important factor that contributes to the elevated publication rate of particular transporters has been expression cloning. In the case of the intestinal Na^+-glucose transporter SGLT1 (SLC5A1), due to its hydrophobic nature and difficulty in purifying, functional expression in *Xenopus laevis* oocytes finally opened the door to successful cloning and molecular characterization (Hediger et al., 1987). This progress led to a substantial increase in SLC study, ultimately leading to structural

Table 1. SLCs Specifically Targeted by FDA-Approved Drugs or Drugs in Active Development

Drug Status	SLC	Common Protein Name	Examples
Approved	SLC5A2	SGLT2	canagliflozin; dapagliflozin
	SLC6A1	GAT1	tiagabine
	SLC6A2	NET	atomoxetine
	SLC6A3	DAT	methylphenidate
	SLC6A4	SERT	fluoxetine; sertraline; citalopram (SSRIs)
	SLC12A1/2	NKCC1/2	furosemide (loop diuretics)
	SLC12A3	NCC	hydrochlorothiazide (thiazide diuretics)
	SLC18A1/2	VMAT1/2	reserpine
	SLC18A2	VMAT2	tetrabenazine
	SLC22 family	OATs	probenecid
	SLC25A4/5/6	ANT1/2/3	clodronate
	SLC29A1	ENT1	dipyridamole
Phase II+ Clinical Trial	SLC5A1 (and SLC5A2)	SGLT1 (and SGLT2)	sotagliflozin
	SLC6A9	GlyT1	bitopertin
	SLC9A3	NHE3	tenapanor
	SLC10A2	IBAT	elobixibat
	SLC22A12	URAT1	lesinurad
	SLC40A1	Ferroportin-1	LY2928057

determination (Faham et al., 2008) and development of an antidiabetic drug class (Abdul-Ghani and DeFronzo, 2014) that acts on its renal homolog SGLT2 (SLC5A2).

Other SLCs became highly studied because they were discovered as targets of existing drugs, with VMAT2 (SLC18A2) representing a specific example of this. Reserpine is a drug that was first marketed in the 1950s as a tranquilizer. The actual mode of action of reserpine was only uncovered 40 years later by scoring for cDNAs conferring the ability to sequester the neurotoxin 1-methyl-4-phenylpyridinium (MPP+) in CHO cells, leading to the discovery of the vesicular amine transporter family SLC18 (Liu et al., 1992). As an example of how the availability of research tools has influenced SLC research, there were no publications at all on SLC30A8 until its first cloning and expression in 2004 (Chimienti et al., 2004). Following this publication and a series of papers genetically linking mutations in this protein with diabetes, in recent years SLC30A8 has become one of the most highly studied SLCs (Rutter and Chimienti, 2015). This spike of activity is clearly displayed in Figure S1A. Even more recently, some SLCs that were previously barely studied have been identified to play key roles in physiology. SLC38A9, an SLC recently found to contribute to amino-acid sensing of mTOR, was ranked 288th in the automated ranking of all time SLC publications (Rebsamen et al., 2015; Wang et al., 2015). With the importance of this SLC now clear and tools available to allow its study, one can anticipate an increase in publication rate for this transporter. As for the bottom-ranked 15% of SLC family members, there are more publications in a PubMed search for "star wars" (72 citations) than on these 70 SLCs combined.

Exploring SLCs as Drug Targets

Regarding SLCs as drug targets, a recent publication suggests 26 different SLCs being the targets of known drugs, or drugs in development (Lin et al., 2015; Rask-Andersen et al., 2013). A closer inspection using more stringent criteria (FDA-approved drugs whose *primary mode of action* is considered to be through action on an SLC) revealed just 12 drug classes. Only 8 of these drug classes are believed to act through selective action at a single SLC, while 4 classes are believed to act non-selectively via two or more SLCs. Only 6 further SLCs are targeted by drugs in active development in phase II clinical trials or beyond (Table 1). Several drugs interact with SLCs in addition to their purported primary target, e.g., amiloride (SLC9A1, NHE1) or sulfasalazine (SLC7A11, xCT), but in such examples, it has not been clearly established that these effects contribute to their clinical pharmacology. The GPCR family, in contrast, is a well-established drug target class that has been the subject of systematic drug discovery efforts for half a century. Even when considering the possibility that GPCRs may be intrinsically more relevant as drug targets, the difference between a few SLC targets and ~100 GPCR targets is likely to reflect a historical bias. Clearly the SLC family is underexplored from the standpoint of drug discovery. Druggability of SLCs appears not to be the main or only barrier here, as the majority of the well-studied SLCs have reported small-molecule inhibitors.

Is it reasonable to expect more SLC-targeting drugs? Around 75% of SLCs are predicted to carry small organic molecules. It has been proposed that proteins that have evolved to bind such species are, on average, privileged with respect to small-molecule druggability (Fauman et al., 2011). Experiences thus far appear to support this prediction, with molecules of high ligand efficiency (an indicator or protein druggability) (Hopkins et al., 2014) being identified in the cases where medicinal chemistry efforts have been attempted against SLCs. Even SLCs that carry only inorganic species have been shown to be druggable, including, for example, the SLC12 family targets of the loop and thiazide diuretics. Thus, SLCs appear to offer the rare potential of an underexplored gene family with high disease relevance and general small-molecule druggability.

SLC Genes and Human Disease

Current thinking in biomedicine and drug discovery contends that human genomics will provide the clues to those genes and proteins of particular relevance to disease and therapy. Accordingly, we looked at all SLC genes that are associated with human disease and counted the number of compounds reported for each (IC_{50} < 10 μM), using OpenPHACTS, a platform that provides a single access to disease, chemical, and target databases (Ratnam et al., 2014; Williams et al., 2012). 76% of SLCs (145 out of 190) with an already identified disease link have no compound associated with them (Figure S2). It is notable how few SLC targets have more than 100 active compounds against them in the database, likely to represent another measure indicative of how

few drug discovery programs have been run against the family. In contrast, the most popular targets of monoamine uptake inhibitors (SLC6A2,3,4) have more than a thousand compounds associated with each, with likely thousands more such compounds in pharmaceutical company collections as a result of extensive drug discovery campaigns against these targets.

Of course, it could be argued that involvement of SLC genes in monogenic disorders is a poor reason to call for drug discovery efforts in the corresponding disease areas, as it appears counterintuitive. Yet such arguments need not be always valid, as there is a fundamental difference between life-long genetic loss of function (LOF) and the titrated, reversible pharmacological blockade of a protein. For instance, LOF mutations in the dopamine transporter SLC6A3 lead to early stage Parkinsonism disease (Kurian et al., 2009), but SLC6A3 is also a principal target of methlyphenidate and in the treatment of psychiatric disorders. Further, LOF mutations in SLC12A3 have been found associated with Gitelman's syndrome, characterized by low blood pressure, and SLC12A3 could be mechanistically linked to the action of thiazides that treat hypertension (Brinkman et al., 2006). Even if we take a more stringent connection to disease by counting only the genetic mutations in the OMIM database (103 different SLCs) (Amberger et al., 2015), it is clear that the "disease" zones of the SLC network are not covered nearly enough by chemical agents.

Why So Little Research Attention Then?

What might have contributed to this apparent anomaly in the distribution of research attention for the SLC gene family, where some members are well studied and so many members not studied at all? First, a unifying nomenclature has been adopted only recently (Hediger et al., 2013, 2004), and as a consequence, common principles and features may have been overlooked. Second, there are a number of technical barriers that may have impeded research in this area. In particular, acquiring competent biological reagents for SLC study can be highly challenging. These are complex integral membrane proteins that are difficult to express and purify and are often poorly detected by typical protocols for mass spectrometry. Accordingly, biochemical, biophysical, and structural biology characterization of SLCs has also been challenging. Indeed, there are so far only three reported human SLC structures (Deng et al., 2014; Gruswitz et al., 2010; Schlessinger et al., 2013a; Deng et al., 2015) (Table S1). Cell-based systems for studying SLC function can likewise be challenging to obtain, as overexpression can cause toxicity (presumably as a result of metabolic perturbation), and loss- or gain-of-function studies can be confounded by endogenous SLCs with overlapping specificities or by compensatory transport or metabolic effects. Even when cell systems with functionally competent SLCs can be obtained, defining their relevant endogenous substrates is not trivial, and establishing screening assays can be difficult. Third, high-quality antibodies are available for only a few SLCs, with the human protein atlas reporting just 45 SLCs for which they have raised reliable antibodies (Uhlen et al., 2015). As a consequence, the current understanding of the subcellular localization of SLCs, crucial for the interpretation of their function, is indeed partial at best. Finally, the transport assays are often challenging, even for those SLCs with known substrates. Artificial lipid vesicles or microinjected frog oocytes, two other useful assay systems, do not necessarily allow for testing function in the context of the regulatory intricacies, and the latter is not always robust enough for large-scale compound screening. In short, despite the post-genomic era, ample evidence for their important physiological role and their druggability, the systematic and parallel structural and functional interrogation of human SLC proteins has not yet been carried out.

Delving into the "Sparse Zones" of Our Knowledge

Here, we argue that an energetic and detailed exploration of the human "SLCome" is warranted because the family comprises one of the largest "sparse zones" of human biology. Indeed, the concept of the rational filling of sparse zones of knowledge is starting to guide strategies in other collaborative efforts (Rolland et al., 2014; Snijder et al., 2014). Furthermore, we argue that the problem should be tackled systematically to capture the efficiencies that come with economies of scale and the learnings that derive from studying related proteins. Finally, we believe that the initial objective of this effort should focus on generating high-quality, enabling reagents (antibodies, purified proteins, cell-based assays, chemical probes, CRISPR-cell lines) and data sets (protein interaction, tissue and sub-cellular distribution).

Such a concerted effort is not only called for but is also timely due to recent technological developments, listed and referenced in Table 2. Such developments cover protein expression, metabolomics, structure determination, gene knockout technologies, and mass spectrometry, as well as assay development and medicinal chemistry, to deliver high-quality chemical tools into the public domain. We listed possible project aims of a concerted campaign, fully aware that such lists are not comprehensive and are meant to spur additional thoughts. There are several examples of successful de-orphanization of SLCs using recently developed technologies (Abplanalp et al., 2013; Caulfield et al., 2008; Iharada et al., 2010; Rebsamen et al., 2015; Wang et al., 2015; Wikoff et al., 2009; Winter et al., 2014).

In broad terms, the strategy to study proteins by family, where experimental methods on one family member may facilitate analysis of the next (Hoglund et al., 2011; Schlessinger et al., 2010), has been highly successful for tackling the structure and chemical tractability of other gene families such as kinases, GPCRs, and proteins involved in the regulation of the epigenome (Barr et al., 2009; Edwards et al., 2009). Importantly, although one would expect similar success applying this approach to the SLC family, there is an additional opportunity that functional inter-relationships among SLCs, on top of phylogeny, may greatly aid in the design of the experimental strategy.

Working Groups of SLCs

It is highly likely that the transport activity of one SLC may affect the activity of others, acting in parallel or in sequential order, in redundant or interdependent function, integrating with the cellular metabolism in various ways (Nigam, 2015; Thiele et al., 2013). If this is the case, there may be several ways to uncover such functionally linked groups, for example, by analyzing co-expression patterns (Huynen et al., 2003; Jordan et al., 2004; Stuart et al., 2003). Proteins acting together are more likely to be co-expressed across tissues and conditions than if they are

Table 2. Approaches to Enable SLC De-orphanization

Objective	Enabling New Technology	References
Expression map of SLCs across the human body, at single-cell and sub-cellular resolution	large-scale RNA-seq; single-cell RNA-seq; expression proteomics and antibody mapping efforts; MALDI imaging mass spectrometry; CyTOF	(Bendall et al., 2011; Clemencon et al., 2015; Cornett et al., 2007; Kim et al., 2014; Mele et al., 2015; Uhlen et al., 2015; Wilhelm et al., 2014)
Human cell lines mutated in individual SLC genes	CRISPR technology; insertional mutagenesis in haploid cells	(Burckstummer et al., 2013; Carette et al., 2009; Doudna and Charpentier, 2014)
Cell lines with multiple SLC gene deletions; cells with minimal SLC repertoire	CRISPR-mediated genomic engineering	(Doudna and Charpentier, 2014; Hsu et al., 2014)
SLC genetic interaction landscape	SLCome- and genome-wide CRISPR inactivation and gain of function libraries; k.o. cells	(Cong et al., 2013; Qi et al., 2013)
Chemical genomics	high-throughput phenotypic screening	(Carette et al., 2009; Reiling et al., 2011; Winter et al., 2014)
SLC interactome	label-free high-throughput AP-MS; BirA-mediated BioID; membrane interaction mapping	(Cox and Mann, 2011; Lambert et al., 2015; Petschnigg et al., 2014; Varjosalo et al., 2013)
Metabolomic data and SLC genetic polymorphisms	genetic association studies; population-wide whole-genome sequencing; rare disease genome sequencing coupled with deep metabolomics	(Shin et al., 2014)
Metabolome-wide transport assays, in dependence of individual SLC gene alteration	high-throughput accurate LC/ GC-mass spectrometry and databases; libraries of metabolites; k.o. cells	(Kell, 2004)
Transport assays using recombinant proteins	proteoliposomes; liposome microarrays; pure solutes, complex body fluids	(Krumpochova et al., 2012; Saliba et al., 2014; Scalise et al., 2013)
High-throughput determination of 3D structure	single-particle cryo-EM; high-throughput crystallization protocols; serial femtosecond crystallography	(Bai et al., 2015; Bartesaghi et al., 2015; Chapman et al., 2011; Moraes et al., 2014; Zeev-Ben-Mordehai et al., 2014; Zhou et al., 2015)
Potent and selective chemical probes for each SLC	better libraries; more accurate screening technologies; assays to assess target engagement and specificity in cells and tissues	(Edwards et al., 2009; Frye, 2010)

functionally independent. SLCs that are consistently identified to be co-expressed and thus represent such putative working groups or functional modules may, in turn, help to shed light on the role of the individual family members. Perhaps these connections can be used to distinguish an underlying overall architecture, which might be suggestive of dependencies and vulnerabilities of the system.

To explore this concept more fully, we analyzed different gene expression data sets of human tissues (Fantom5, Illumina body map, and the "32 tissues"; Forrest et al., 2014; Parkinson et al., 2011; Uhlen et al., 2015) and derived a global and high-confidence survey of patterns of co-expression across SLC genes. These patterns for co-expression were analyzed to identify SLCs that are frequently and consistently co-expressed. Co-expression relationships were ranked based on the combined p values of the correlations in the three independent data sets used. For visualization convenience, we chose to display only the top 2,500 co-expressed SLCs observed in at least two data sets. We found at least five major clusters and several smaller ones, perhaps representing fundamental functional relationships (Figure 2A). The edges were colored according to the tissue in which two connected SLCs are most highly, but not necessarily exclusively, expressed. We found that the clusters correspond to individual tissues (kidney, liver, brain, testis, leukocytes). Interestingly, kidney and liver seem to share the highest number of co-expressed SLCs despite their different germ-layer origin. A more fine-grained tissue annotation shows that the kidney/liver intersection harbors the SLCs whose co-expression is highest in intestinal tissue (Figure S3A). This similarity between kidney and liver co-expression is specific for SLCs, as a recently published genome-wide tissue expression comparison revealed a considerably larger "distance" between these organs (Mele et al., 2015).

The network displayed does not automatically reveal all textbook cases of co-expression. For example, expression of SLC26A4 (iodide transporter, pendrin) and SLC5A5 (sodium iodide co-transporter) is well known to be coordinately expressed in thyroid tissue. While a significant level of co-expression is observed in the thyroid, several tissues either express one or the other, suggesting that they may not always be obligatory partners and may have independent functions. Accordingly, the correlation, although significant, did not reach the top

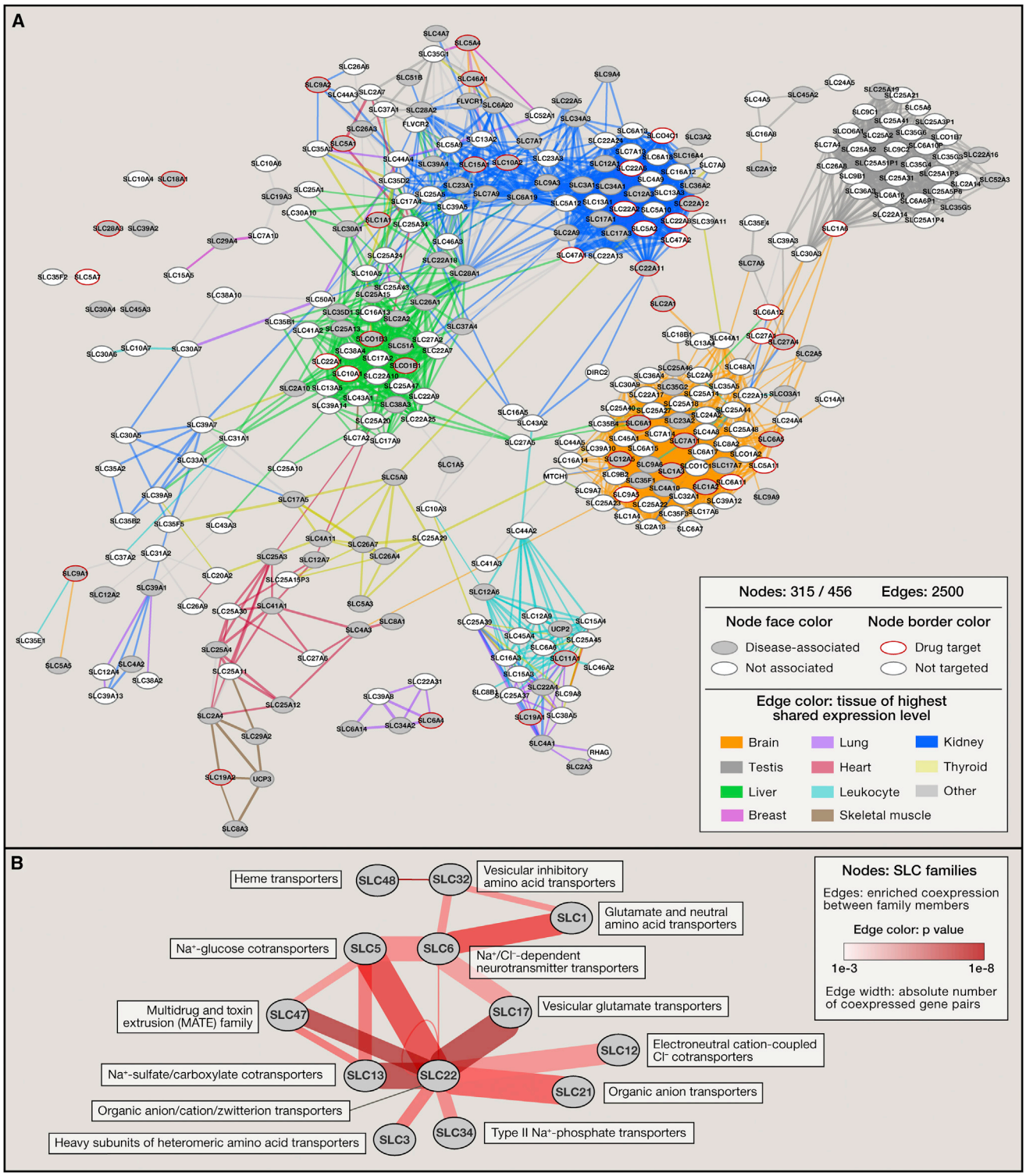


Figure 2. SLCs Are Expressed in Robust Tissue-Dependent Modules

(A) Network visualization of SLC co-expression. Nodes in the network represent SLCs. Edges between nodes correspond to significant correlations consistently retrieved in three independent expression data sets from healthy human tissues. Only the top 2,500 most significant edges are shown, based on combined p values of the three independent correlations. Gray nodes indicate SLCs with at least one disease association, and red node outlines indicate the presence of at

(legend continued on next page)

2,500. To allow the interrogation of the full data set beyond what can be reasonably visualized in a single network, we include an extended list of SLC co-expression pairs across all tissues (Table S2).

Not all SLCs are represented in the network because their expression does not correlate strongly enough with any given other to be among the top 2,500 that we chose to represent for visualization. Of these SLCs, some 48 appear to be expressed in all tissues and may thus represent a "core" of housekeeping functions (Table S3). Membership to this group may make some of them attractive to study for pharmacokinetic considerations.

We then looked at the distribution of disease-associated SLCs across the network (dark gray filled nodes, Figure 2A). All clusters (except the testis cluster) contain several positive SLCs, confirming that SLC gene functions bear important pathophysiological implications across many tissues and processes. The SLCs for which high-affinity chemical agents have been developed are marked; their distribution was considerably less even (red halos, Figure 2A). At least three cluster regions seem relatively sparse in terms of drugs: heart and skeletal muscle, leukocytes, and the intersection of liver and kidney. Perhaps these regions merit more attention in the future.

The SLC families do not appear to group in clusters or tissues (i.e., most SLC families appear distributed over the different tissues), but there is a non-random pairing of co-expression between different SLC families whose pattern likely reflects metabolic/biochemical dependencies (Figure 2B). For instance, strong interaction between the SLC13 and SLC22 families is likely to reflect an integration of energy and homeostatic regulation of intermediate metabolism, particularly the Krebs cycle. Enrichment in the interactions between families SLC5 (glucose reuptake), SLC13 (citrate/dicarboxylate reuptake), and SLC47 (toxin/xenobiotic secretion) might be also explained by the role of some of their members in kidney, where a coordinated transport of their cargos is required. Furthermore, the sodium and chloride symporter family SLC6, which transports monoamine neurotransmitters and amino acid neurotransmitters, is heavily linked with glutamate/neutral amino acid transporters of the SLC1 family. This link suggests a connection at both a metabolic and physiological level, especially important in brain tissue.

The robustness of SLC co-expression patterns across different large-scale data sets was very high and clearly exceeded, for example, that of protein kinases (Figure S3B). Possibly, SLC function has a particularly high degree of interdependence reflective of the integrative nature of metabolism required for homeostatic stability. The resulting co-expression networks are likely to be reinforced by the integration with environmental parameters.

The patterns of SLC co-expression may reflect normal cell function; when we compared co-expression in different cancer cell lines, we observed massive changes, corresponding to a general loss of structural organization in the network. The network of SLC co-expression in cancer is not robust, i.e., it is not as reproducible across data sets, and it shows considerably less clustering (Figure S3C). The degree by which cancer perturbs the SLC co-expression pattern is much higher than the differences between normal tissues and cancer cell lines observed with protein kinases.

Could loss of these "healthy" SLC co-expression patterns be a good marker for the loss of homeostasis in certain diseases? This analysis may suggest that there is a SLC regulatory circuitry that may be crucial to medical and pharmacological considerations and that might assist strategic choices in the effort to fill the SLC knowledge gap. Armed with this knowledge, redundancy and dependency are not annoying impediments of the large SLC group but, rather, exploitable features.

Conclusions

In summary, SLCs are particularly understudied and fascinating proteins, vital for correct cellular function by controlling the correct import and export of the molecules of life across membranes. They are important in disease and in the action and transport of drugs. A broad attack on their structure, expression, regulation, chemical structure-activity relationships, and functional characterization in terms of transport and signaling is warranted. The study of their regulation and interdependencies should be particularly fascinating, as the functional target may be not only a single protein but the vulnerability within the functional network, perhaps involving ATP-dependent efflux transporters of the ABC family as well. A full-force effort to study the "SLCome" would open the doors to the interface between human health and metabolism, nutrition, and the environment. The large and important family of SLCs should be neglected no longer.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.07.022.

ACKNOWLEDGMENTS

We thank all of the participants of the SLC workshop in Vienna, November 6 and 7, 2014 at CeMM and various members of the Superti-Furga laboratory for discussions and H.E. Susan le Jeune d'Allegeershecque and the British Embassy for graceful hosting. The work was funded by the Austrian Academy of Sciences (G.S.-F., A.C.R., B.S.), ERC grant agreement number 250179 (G.S.-F), the Swiss National Science Foundation (project number P300P3_147897 to B.S.), the Marie Curie Actions International Fellowship Program (IFP) TransCure (G.G.), and the Swiss National Center of Competence in Research (NCCR TransCure). D.H. is an employee and shareholder of Pfizer, Inc. G.S.-F. is a founder and shareholder of Haplogen, GmbH. The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Eshelman Institute for Innovation, Genome Canada, Janssen, Merck & Co., Novartis Pharma AG, the Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and the Wellcome Trust.

least one interacting small compound with an IC_{50} below 10 μM (OpenPHACTS). Edges are colored according to the tissue in which the two connected SLCs share the highest expression (highest mean rank; Illumina Body Map data set), as indicated in the legend.

(B) SLC family co-expression enrichment network. Nodes in the network represent SLC families. Edges correspond to statistically enriched co-expression between members of the connected SLC families, as calculated by a hypergeometric test. Edge color relates to the significance of the enrichment and edge width is proportional to the number of co-expressed SLC pairs (see legend).

REFERENCES

Abdul-Ghani, M.A., and DeFronzo, R.A. (2014). Lowering plasma glucose concentration by inhibiting renal sodium-glucose cotransport. J. Intern. Med. *276*, 352–363.

Abplanalp, J., Laczko, E., Philp, N.J., Neidhardt, J., Zuercher, J., Braun, P., Schorderet, D.F., Munier, F.L., Verrey, F., Berger, W., et al. (2013). The cataract and glucosuria associated monocarboxylate transporter MCT12 is a new creatine transporter. Hum. Mol. Genet. *22*, 3218–3226.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. *43*, D789–D798.

Bai, X.C., McMullan, G., and Scheres, S.H. (2015). How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. *40*, 49–57.

Bailey, D.G. (2010). Fruit juice inhibition of uptake transport: a new type of food-drug interaction. Br. J. Clin. Pharmacol. *70*, 645–655.

Barr, A.J., Ugochukwu, E., Lee, W.H., King, O.N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N.A., Muller, S., and Knapp, S. (2009). Large-scale structural analysis of the classical human protein tyrosine phosphatome. Cell *136*, 352–363.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L., and Subramaniam, S. (2015). Electron microscopy. 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science *348*, 1147–1151.

Bendall, S.C., Simonds, E.F., Qiu, P., El-ad, D.A., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., and Ornatsky, O.I. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

Brinkman, R.R., Dube, M.P., Rouleau, G.A., Orr, A.C., and Samuels, M.E. (2006). Human monogenic disorders - a source of novel drug targets. Nat. Rev. Genet. *7*, 249–260.

Bruford, E.A., Lush, M.J., Wright, M.W., Sneddon, T.P., Povey, S., and Birney, E. (2008). The HGNC Database in 2008: a resource for the human genome. Nucleic Acids Res. *36*, D445–D448.

Burckstummer, T., Banning, C., Hainzl, P., Schobesberger, R., Kerzendorfer, C., Pauler, F.M., Chen, D., Them, N., Schischlik, F., Rebsamen, M., et al. (2013). A reversible gene trap collection empowers haploid genetics in human cells. Nat. Methods *10*, 965–971.

Carette, J.E., Guimaraes, C.P., Varadarajan, M., Park, A.S., Wuethrich, I., Godarova, A., Kotecki, M., Cochran, B.H., Spooner, E., Ploegh, H.L., et al. (2009). Haploid genetic screens in human cells identify host factors used by pathogens. Science *326*, 1231–1235.

Caulfield, M.J., Munroe, P.B., O'Neill, D., Witkowska, K., Charchar, F.J., Doblado, M., Evans, S., Eyheramendy, S., Onipinla, A., Howard, P., et al. (2008). SLC2A9 is a high-capacity urate transporter in humans. PLoS Med. *5*, e197.

Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., DePonte, D.P., Weierstall, U., et al. (2011). Femtosecond X-ray protein nanocrystallography. Nature *470*, 73–77.

Chimienti, F., Devergnas, S., Favier, A., and Seve, M. (2004). Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. Diabetes *53*, 2330–2337.

Clemencon, B., Fine, M., Schneider, P., and Hediger, M.A. (2015). Rapid method to express and purify human membrane protein using the Xenopus oocyte system for functional and low-resolution structural analysis. Methods Enzymol. *556*, 241–265.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., and Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cornett, D.S., Reyzer, M.L., Chaurand, P., and Caprioli, R.M. (2007). MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. Nat. Methods *4*, 828–833.

Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem. *80*, 273–299.

Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M., and Yan, N. (2014). Crystal structure of the human glucose transporter GLUT1. Nature *510*, 121–125.

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S., et al. (2015). Molecular basis of ligand recognition and transport by glucose transporters. Nature. Published online July 15, 2015. http://dx.doi.org/10.1038/nature14655.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Edwards, A.M., Bountra, C., Kerr, D.J., and Willson, T.M. (2009). Open access chemical and clinical probes to support drug discovery. Nat. Chem. Biol. *5*, 436–440.

Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., and Yu, F.H. (2011). Too many roads not taken. Nature *470*, 163–165.

Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A., Hirayama, B.A., Wright, E.M., and Abramson, J. (2008). The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. Science *321*, 810–814.

Fauman, E.B., Rai, B.K., and Huang, E.S. (2011). Structure-based druggability assessment–identifying suitable targets for small molecule therapeutics. Curr. Opin. Chem. Biol. *15*, 463–468.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassman, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Frye, S.V. (2010). The art of the chemical probe. Nat. Chem. Biol. *6*, 159–161.

Giacomini, K.M., Huang, S.M., Tweedie, D.J., Benet, L.Z., Brouwer, K.L., Chu, X., Dahlin, A., Evers, R., Fischer, V., Hillgren, K.M., et al. (2010). Membrane transporters in drug development. Nat. Rev. Drug Discov. *9*, 215–236.

Giacomini, K.M., Balimane, P.V., Cho, S.K., Eadon, M., Edeki, T., Hillgren, K.M., Huang, S.M., Sugiyama, Y., Weitz, D., Wen, Y., et al. (2013). International Transporter Consortium commentary on clinically important transporter polymorphisms. Clin. Pharmacol. Ther. *94*, 23–26.

Gruswitz, F., Chaudhary, S., Ho, J.D., Schlessinger, A., Pezeshki, B., Ho, C.M., Sali, A., Westhoff, C.M., and Stroud, R.M. (2010). Function of human Rh based on structure of RhCG at 2.1 A. Proc. Natl. Acad. Sci. USA *107*, 9638–9643.

Hediger, M.A., Coady, M.J., Ikeda, T.S., and Wright, E.M. (1987). Expression cloning and cDNA sequencing of the Na+/glucose co-transporter. Nature *330*, 379–381.

Hediger, M.A., Romero, M.F., Peng, J.B., Rolfs, A., Takanaga, H., and Bruford, E.A. (2004). The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. Pflugers Arch. *447*, 465–468.

Hediger, M.A., Clemencon, B., Burrier, R.E., and Bruford, E.A. (2013). The ABCs of membrane transporters in health and disease (SLC series): introduction. Mol. Aspects Med. *34*, 95–107.

Hoglund, P.J., Nordstrom, K.J., Schioth, H.B., and Fredriksson, R. (2011). The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. Mol. Biol. Evol. *28*, 1531–1541.

Hopkins, A.L., Keseru, G.M., Leeson, P.D., Rees, D.C., and Reynolds, C.H. (2014). The role of ligand efficiency metrics in drug discovery. Nat. Rev. Drug Discov. *13*, 105–121.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Huynen, M.A., Snel, B., von Mering, C., and Bork, P. (2003). Function prediction and protein networks. Curr. Opin. Cell Biol. *15*, 191–198.

Iharada, M., Miyaji, T., Fujimoto, T., Hiasa, M., Anzai, N., Omote, H., and Moriyama, Y. (2010). Type 1 sodium-dependent phosphate transporter (SLC17A1 Protein) is a Cl(-)-dependent urate exporter. J. Biol. Chem. *285*, 26107–26113.

Jordan, I.K., Marino-Ramirez, L., Wolf, Y.I., and Koonin, E.V. (2004). Conservation and coevolution in the scale-free human gene coexpression network. Mol. Biol. Evol. *21*, 2058–2070.

Kell, D.B. (2004). Metabolomics and systems biology: making sense of the soup. Curr. Opin. Microbiol. *7*, 296–307.

Kell, D.B., and Oliver, S.G. (2014). How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion. Front. Pharmacol. *5*, 231.

Kell, D.B., Dobson, P.D., and Oliver, S.G. (2011). Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only. Drug Discov. Today *16*, 704–714.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat. Biotechnol. *33*, 306–312.

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. Nature *509*, 575–581.

Koberlin, M.S., Snijder, B., Heinz, L.X., Baumann, C.L., Fauster, A., Vladimer, G.I., Gavin, A.C., and Superti-Furga, G. (2015). A Conserved Circular Network of Coregulated Lipids Modulates Innate Immune Responses. Cell *162*, 170–183.

Krumpochova, P., Sapthu, S., Brouwers, J.F., de Haas, M., de Vos, R., Borst, P., and van de Wetering, K. (2012). Transportomics: screening for substrates of ABC transporters in body fluids using vesicular transport assays. FASEB J. *26*, 738–747.

Kurian, M.A., Zhen, J., Cheng, S.Y., Li, Y., Mordekar, S.R., Jardine, P., Morgan, N.V., Meyer, E., Tee, L., Pasha, S., et al. (2009). Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. J. Clin. Invest. *119*, 1595–1603.

Lambert, J.P., Tucholska, M., Go, C., Knight, J.D., and Gingras, A.C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. J. Proteomics *118*, 81–94.

Lin, L., Yee, S.W., Kim, R.B., and Giacomini, K.M. (2015). SLC transporters as therapeutic targets: emerging opportunities. Nat. Rev. Drug Discov. Published online June 26, 2015. http://dx.doi.org/10.1038/nrd4626.

Liu, Y., Peter, D., Roghani, A., Schuldiner, S., Prive, G.G., Eisenberg, D., Brecha, N., and Edwards, R.H. (1992). A cDNA that suppresses MPP+ toxicity encodes a vesicular amine transporter. Cell *70*, 539–551.

Mele, 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., et al. (2015). Human genomics. The human transcriptome across tissues and individuals. Science *348*, 660–665.

Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., and Stewart, P.D. (2014). Membrane protein structure determination - the next generation. Biochim. Biophys. Acta *1838*, 78–87.

Mulkidjanian, A.Y., Galperin, M.Y., and Koonin, E.V. (2009). Co-evolution of primordial membranes and membrane proteins. Trends Biochem. Sci. *34*, 206–215.

Nigam, S.K. (2015). What do drug transporters really do? Nat. Rev. Drug Discov. *14*, 29–44.

Parkinson, H., Sarkans, U., Kolesnikov, N., Abeygunawardena, N., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Holloway, E., et al. (2011). ArrayExpress update–an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic Acids Res. *39*, D1002–D1004.

Petschnigg, J., Groisman, B., Kotlyar, M., Taipale, M., Zheng, Y., Kurat, C.F., Sayad, A., Sierra, J.R., Mattiazzi Usaj, M., Snider, J., et al. (2014). The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. Nat. Methods *11*, 585–592.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Rask-Andersen, M., Masuram, S., Fredriksson, R., and Schioth, H.B. (2013). Solute carriers as drug targets: current use, clinical trials and prospective. Mol. Aspects Med. *34*, 702–710.

Ratnam, J., Zdrazil, B., Digles, D., Cuadrado-Rodriguez, E., Neefs, J.M., Tipney, H., Siebes, R., Waagmeester, A., Bradley, G., Chau, C.H., et al. (2014). The application of the open pharmacological concepts triple store (open PHACTS) to support drug discovery research. PLoS ONE *9*, e115460.

Rebsamen, M., Pochini, L., Stasyk, T., de Araujo, M.E., Galluccio, M., Kandasamy, R.K., Snijder, B., Fauster, A., Rudashevskaya, E.L., Bruckner, M., et al. (2015). SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. Nature *519*, 477–481.

Reiling, J.H., Clish, C.B., Carette, J.E., Varadarajan, M., Brummelkamp, T.R., and Sabatini, D.M. (2011). A haploid genetic screen identifies the major facilitator domain containing 2A (MFSD2A) transporter as a key mediator in the response to tunicamycin. Proc. Natl. Acad. Sci. USA *108*, 11756–11765.

Rolland, T., Tasan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Rutter, G.A., and Chimienti, F. (2015). SLC30A8 mutations in type 2 diabetes. Diabetologia *58*, 31–36.

Saliba, A.E., Vonkova, I., Ceschia, S., Findlay, G.M., Maeda, K., Tischer, C., Deghou, S., van Noort, V., Bork, P., Pawson, T., et al. (2014). A quantitative liposome microarray to systematically characterize protein-lipid interactions. Nat. Methods *11*, 47–50.

Scalise, M., Pochini, L., Giangregorio, N., Tonazzi, A., and Indiveri, C. (2013). Proteoliposomes as tool for assaying membrane transporter functions and interactions with xenobiotics. Pharmaceutics *5*, 472–497.

Schlessinger, A., Matsson, P., Shima, J.E., Pieper, U., Yee, S.W., Kelly, L., Apeltsin, L., Stroud, R.M., Ferrin, T.E., Giacomini, K.M., et al. (2010). Comparison of human solute carriers. Protein Sci. *19*, 412–428.

Schlessinger, A., Khuri, N., Giacomini, K.M., and Sali, A. (2013a). Molecular modeling and ligand docking for solute carrier (SLC) transporters. Curr. Top. Med. Chem. *13*, 843–856.

Schlessinger, A., Yee, S.W., Sali, A., and Giacomini, K.M. (2013b). SLC classification: an update. Clin. Pharmacol. Ther. *94*, 19–23.

Schrum, J.P., Zhu, T.F., and Szostak, J.W. (2010). The origins of cellular life. Cold Spring Harb. Perspect. Biol. *2*, a002212.

Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al. (2014). An atlas of genetic influences on human blood metabolites. Nat. Genet. *46*, 543–550.

Snijder, B., Kandasamy, R.K., and Superti-Furga, G. (2014). Toward effective sharing of high-dimensional immunology data. Nat. Biotechnol. *32*, 755–759.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249–255.

Thiele, I., Swainston, N., Fleming, R.M., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat. Biotechnol. *31*, 419–425.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Varjosalo, M., Sacco, R., Stukalov, A., van Drogen, A., Planyavsky, M., Hauri, S., Aebersold, R., Bennett, K.L., Colinge, J., Gstaiger, M., et al. (2013). Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. Nat. Methods *10*, 307–314.

Wang, S., Tsun, Z.Y., Wolfson, R.L., Shen, K., Wyant, G.A., Plovanich, M.E., Yuan, E.D., Jones, T.D., Chantranupong, L., Comb, W., et al. (2015). Metabolism. Lysosomal amino acid transporter SLC38A9 signals arginine sufficiency to mTORC1. Science *347*, 188–194.

Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G. (2009). Metabolomics analysis reveals large effects of gut

microflora on mammalian blood metabolites. Proc. Natl. Acad. Sci. USA *106*, 3698–3703.

Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature *509*, 582–587.

Williams, A.J., Harland, L., Groth, P., Pettifer, S., Chichester, C., Willighagen, E.L., Evelo, C.T., Blomberg, N., Ecker, G., Goble, C., et al. (2012). Open PHACTS: semantic interoperability for drug discovery. Drug Discov. Today *17*, 1188–1198.

Williams, A.L., Jacobs, S.B., Moreno-Macias, H., Huerta-Chagoya, A., Churchhouse, C., Marquez-Luna, C., Garcia-Ortiz, H., Gomez-Vazquez, M.J., Burtt, N.P., Aguilar-Salinas, C.A., et al. (2014). Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature *506*, 97–101.

Winter, G.E., Radic, B., Mayor-Ruiz, C., Blomen, V.A., Trefzer, C., Kandasamy, R.K., Huber, K.V., Gridling, M., Chen, D., Klampfl, T., et al. (2014). The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. Nat. Chem. Biol. *10*, 768–773.

Zeev-Ben-Mordehai, T., Vasishtan, D., Siebert, C.A., Whittle, C., and Grunewald, K. (2014). Extracellular vesicles: a platform for the structure determination of membrane proteins by Cryo-EM. Structure *22*, 1687–1692.

Zhang, Q., Zhang, Y., Diamond, S., Boer, J., Harris, J.J., Li, Y., Rupar, M., Behshad, E., Gardiner, C., Collier, P., et al. (2014). The Janus kinase 2 inhibitor fedratinib inhibits thiamine uptake: a putative mechanism for the onset of Wernicke's encephalopathy. Drug Metab. Dispos. *42*, 1656–1662.

Zhou, Q., Huang, X., Sun, S., Li, X., Wang, H.W., and Sui, S.F. (2015). Cryo-EM structure of SNAP-SNARE assembly in 20S particle. Cell Res. *25*, 551–560.

Supplemental Figures

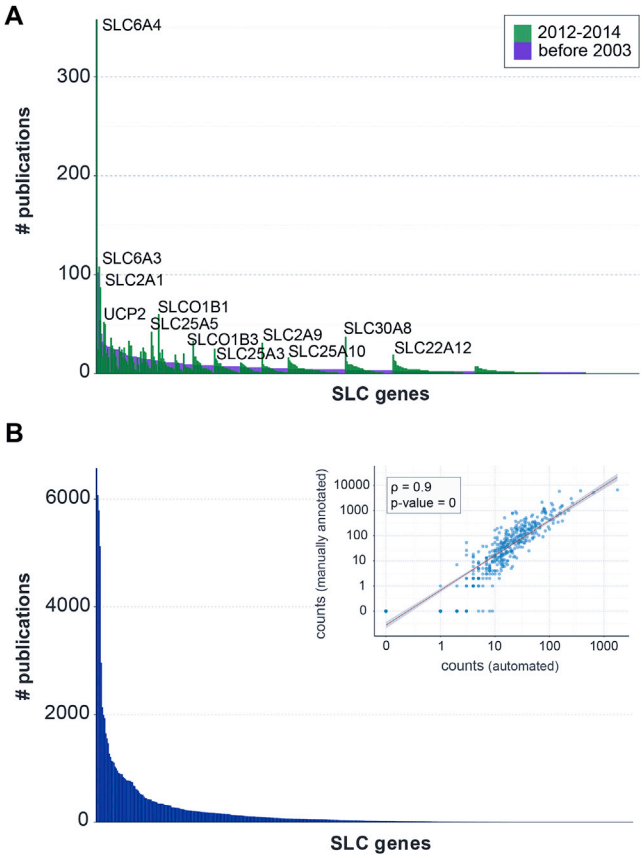


Figure S1. Publication Bias on SLCs

(A) Number of publications per SLC gene before 2003 (purple) and between 2012 and 2014 (green) are displayed in descending order according to the before 2003 count. Top and upcoming SLCs are labeled.

(B) Same as Figure 1B, but publications counts were retrieved by manual querying of PubMed abstracts including all gene names and synonyms for each SLC. Insert shows high concordance between the automated (x axis) and the manually queried (y axis) publication counts per SLC.

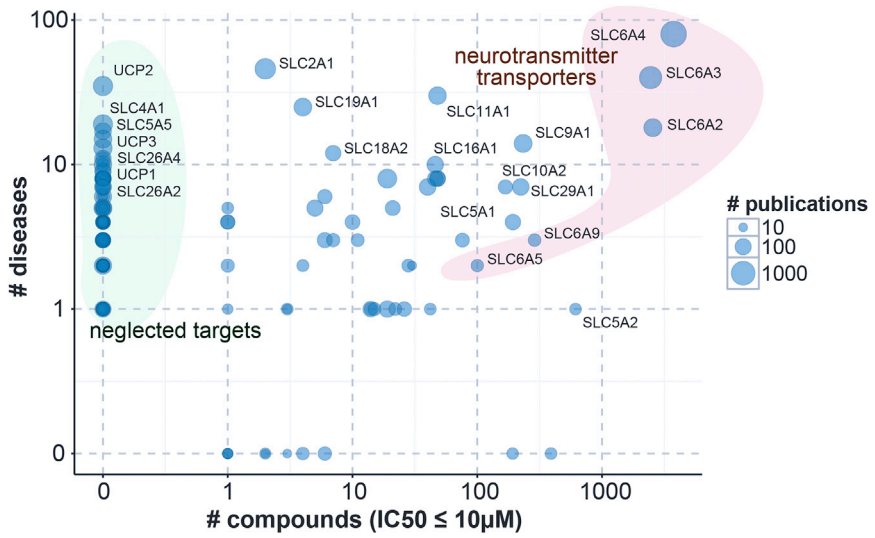


Figure S2. Few Disease-Associated SLCs Are Targeted by Small Compounds

Number of disease associations (y axis) and small compound interactions (IC_{50} smaller than 10μM; x axis) are shown for each SLC, in log scale. All data were retrieved from OpenPHACTS. SLCs with neither disease associations nor interacting small compounds are excluded. Dot size indicates number of publications (see legend). Neurotransmitter transporters (SLC family 6; red) and neglected targets (green) are highlighted.

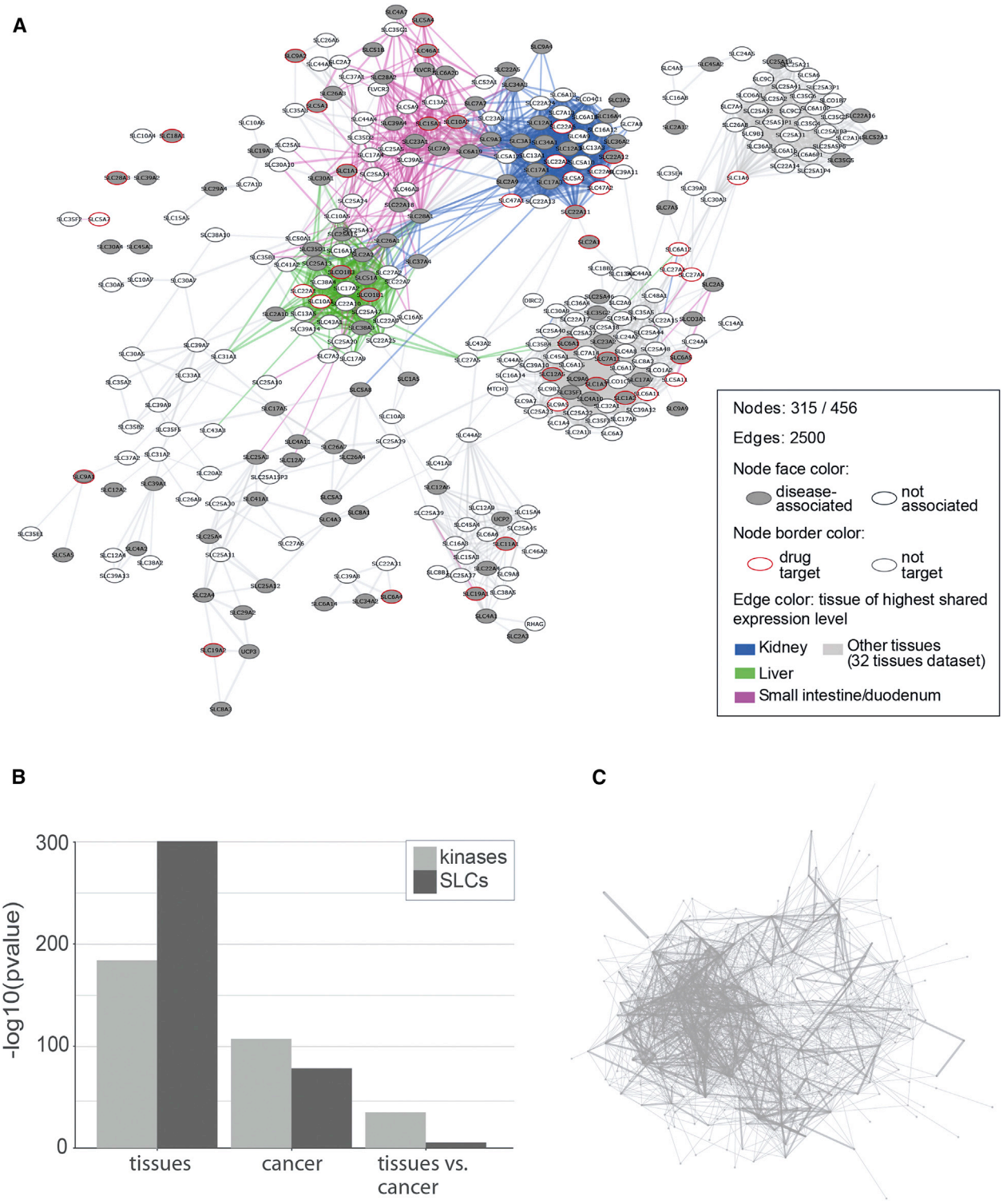


Figure S3. SLCs Are Expressed in Robust Tissue-Dependent Modules

(A) Same as Figure 2A, but edge colors are annotated with the 32 tissues dataset instead of the Illumina Body Map dataset. Only selected tissues are highlighted: liver (green) and kidney (blue) as in Figure 2A, together with small intestine and duodenum (pink). Other tissues are colored in gray.

(B) Significance of the overlap of top 1000 correlations between tissue datasets (32 tissues, Illumina Body Map, Fantom5 – tissue samples), between cancer datasets (CCLE, Cosmic, and Cancer Cell Lines (Klijn et al., 2015) and between tissue and cancer datasets. P-values were calculated by a hypergeometric test.

(C) Same as in Figure 2A, but SLC co-expression is analyzed in cancer expression datasets (CCLE, Cosmic, and Cancer Cell Lines).

2.3 Interlude I

The lack of knowledge regarding the transport mechanism of most drugs and drug-like compounds combined with the understudied nature of Solute carriers (SLCs) opens the door to the potential discovery of new drug-transporter relationships that can help to decipher the mechanism of action of many therapeutic agents. To this end, the large-scale pharmacogenomics datasets on panels of cancer cell lines developed in the last years constitute a rich source of data that can be exploited using computational methods in order to identify transporters that affect drug response.

In the next publication, we aimed at prioritizing SLC (and ABC) drug relationships by building predictive models of drug sensitivity using the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, one of the largest and most comprehensive to date, which includes 1,000 molecularly annotated cancer cell lines together with their response to 265 anti-cancer drugs. To get an idea on the current knowledge of drug transport mediated by SLCs and ABCs, we first surveyed public pharmacology resources and relevant publications, and obtained a connected network that involved 100 transporters (approximately a fifth of the total) and 500 compounds. We then characterized the expression of SLC and ABC genes across the GDSC cancer cell line panel, roughly distinguishing similarly sized sets of specific and commonly expressed transporters. Finally, we applied regularized linear regression methods (LASSO and Elastic Net) to predict drug response based on SLC and ABC molecular data only. The most predictive models identified both known and uncharacterized associations for a number of compounds, providing a prioritization means of potentially pharmacologically relevant interactions.

2.4 Manuscript #2

César-Razquin A, Girardi E, Yang M, Brehme M, Saez-Rodriguez J & Superti-Furga G (2018) In silico Prioritization of Transporter-Drug Relationships From Drug Sensitivity Screens. Front Pharmacol 9: 1011 doi: 10.3389/fphar.2018.01011

Detailed author contribution: A. César-Razquin performed most of the analyses, designed all figures and co-wrote the manuscript; E.G. supervised the project and co-wrote the manuscript; M.Y. performed the LASSO analysis; M.B., J.S-R, and G.S-F provided scientific insight and project supervision. All co-authors read and commented on the manuscript.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 07 September 2018
doi: 10.3389/fphar.2018.01011

In silico Prioritization of Transporter–Drug Relationships From Drug Sensitivity Screens

Adrián César-Razquin[1], Enrico Girardi[1], Mi Yang[2,3], Marc Brehme[3,4], Julio Saez-Rodriguez[3,5†] and Giulio Superti-Furga[1,6]*

[1] CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences,Vienna, Austria, [2] Faculty of Biosciences, Heidelberg University, Heidelberg, Germany, [3] RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC-COMBINE), Aachen, Germany, [4] CBmed – Center for Biomarker Research in Medicine GmbH, Graz, Austria, [5] European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom, [6] Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:
Yurong Lai,
Gilead, United States

Reviewed by:
Zhihao Liu,
University of Illinois at Chicago, United States
Peng Hsiao,
Ferring Research Institute, Inc., United States

***Correspondence:**
Giulio Superti-Furga
gsuperti@cemm.oeaw.ac.at

†Present address:
Julio Saez-Rodriguez,
Faculty of Medicine, Institute for Computational Biomedicine, Heidelberg University, Heidelberg, Germany

Specialty section:
This article was submitted to Drug Metabolism and Transport, a section of the journal Frontiers in Pharmacology

Received: *05 July 2018*
Accepted: *20 August 2018*
Published: *07 September 2018*

Citation:
César-Razquin A, Girardi E, Yang M, Brehme M, Saez-Rodriguez J and Superti-Furga G (2018) In silico Prioritization of Transporter–Drug Relationships From Drug Sensitivity Screens. Front. Pharmacol. 9:1011. doi: 10.3389/fphar.2018.01011

The interplay between drugs and cell metabolism is a key factor in determining both compound potency and toxicity. In particular, how and to what extent transmembrane transporters affect drug uptake and disposition is currently only partially understood. Most transporter proteins belong to two protein families: the ATP-Binding Cassette (ABC) transporter family, whose members are often involved in xenobiotic efflux and drug resistance, and the large and heterogeneous family of solute carriers (SLCs). We recently argued that SLCs are collectively a rather neglected gene group, with most of its members still poorly characterized, and thus likely to include many yet-to-be-discovered associations with drugs. We searched publicly available resources and literature to define the currently known set of drugs transported by ABCs or SLCs, which involved ∼500 drugs and more than 100 transporters. In order to extend this set, we then mined the largest publicly available pharmacogenomics dataset, which involves approximately 1,000 molecularly annotated cancer cell lines and their response to 265 anti-cancer compounds, and used regularized linear regression models (Elastic Net, LASSO) to predict drug responses based on SLC and ABC data (expression levels, SNVs, CNVs). The most predictive models included both known and previously unidentified associations between drugs and transporters. To our knowledge, this represents the first application of regularized linear regression to this set of genes, providing an extensive prioritization of potentially pharmacologically interesting interactions.

Keywords: solute carriers, ABC transporters, drug sensitivity and resistance, drug transport, regularized linear regression

INTRODUCTION

The role of cellular metabolism in determining the potency and distribution of drugs is increasingly recognized (Zhao et al., 2013). Along with the enzymes involved in actual xenobiotic transformation, such as members of the cytochrome and transferases families, a critical role is played by transmembrane transporters, which directly affect both the uptake and the excretion of drugs and their metabolites (Zhou et al., 2017). Among transmembrane transporters, two

large families have been described: the family of ATP-binding cassette (ABC) transporters and the family of solute carriers (SLCs) (Hediger et al., 2013). ABC transporters are pumps powered by the hydrolysis of ATP and show a remarkable broad range of substrates, including lipids, secondary metabolites, and xenobiotics. Members of this family, such as the ABCB/MDR and ABCC/MRP proteins, have been associated with resistance to a large number of structurally diverse compounds in cancer cells (Fletcher et al., 2010). SLCs are secondary transporters involved in uptake or efflux of metabolites and other chemical matter (Cesar-Razquin et al., 2015). At more than 400 members and counting, SLCs represent the second largest family of membrane proteins and comprise uniporters, symporters and antiporters, further grouped into subfamilies based on sequence similarity (Hoglund et al., 2011). Among the reported SLC substrates are all major building blocks of the cell, such as nucleic acids, sugars, lipids, and amino acids as well as vitamins, metals, and other ions (Hediger et al., 2013). Despite the critical processes mediated by these proteins, a large portion of SLCs is still poorly characterized and, in several cases, lacks any associations with a substrate (Cesar-Razquin et al., 2015). Importantly, several members of the SLCO (also known as Organic Anion Transporter Proteins or OATPs) and SLC22 families (including the group of organic cation transporters or OCTs, organic zwitterion/cation transporters or OCTNs and organic anion transporters or OATs) have been found to play prominent roles in the uptake and excretion of drugs, especially in the liver and kidneys (Hagenbuch and Stieger, 2013). Several other cases of SLCs mediating the uptake of drugs have been reported, such as in the case of methotrexate and related anti-folate drugs with the folate transporter SLC19A1 (Zhao et al., 2011) or the anti-cancer drug YM155/sepantronium bromide and the orphan transporter SLC35F2 (Winter et al., 2014). Indeed, whether carrier-mediated uptake is the rule or rather the exception is still a matter of discussion (Dobson and Kell, 2008; Sugano et al., 2010). Due to the understudied nature of transporters and SLCs in particular, we can nonetheless expect that several other associations between drugs and transporters, involving direct transport or indirect effects, remain to be discovered and could provide novel insights into the pharmacokinetics of drugs and drug-like compounds.

Analysis of basal gene expression and genomic features in combination with drug sensitivity data allows the identification of molecular markers that render cells both sensitive and resistant to specific drugs. Such a pharmacogenomic analysis represents a powerful method to prioritize *in silico* gene-compound associations. Different statistical and machine learning (ML) strategies have been used in the past to confirm known as well as to identify novel drug–gene associations, although generally in a genome-wide context (Iorio et al., 2016). For our study, we mined the "Genomics of Drug Sensitivity in Cancer" (GDSC) dataset (Iorio et al., 2016) which contains drug sensitivity data to a set of 265 anti-cancer compounds over ∼1,000 molecularly annotated cancer cell lines, in order to explore drug relationships exclusively involving transporters (SLCs and ABCs). To such end, we used regularized linear regression (Elastic Net, LASSO) to generate predictive models from which to extract cooperative sensitivity and resistance drug–transporter relationships, in what represents, to our knowledge, the first work applying this type of analysis to this group of genes.

MATERIALS AND METHODS

Data

Solute carriers and ABC genes were considered as in (Cesar-Razquin et al., 2015). Known drug transport cases involving SLC and ABC proteins were obtained from four main repositories as of September 2017: DrugBank (Law et al., 2014), The IUPHAR/BPS Guide to PHARMACOLOGY (Alexander et al., 2015), KEGG: Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto, 2000), and UCSF-FDA TransPortal (Morrissey et al., 2012). These data were complemented with various other cases found in the literature (Sprowl and Sparreboom, 2014; Winter et al., 2014; Nigam, 2015; Radic-Sarikas et al., 2017). Source files were parsed using custom python scripts, and all entries were manually curated, merged together and redundancies eliminated. The final compound list was searched against PubChem (Kim et al., 2016) in order to systematize names. A list of FDA-approved drugs was obtained from the organization's website. Network visualization was done using Cytoscape (Shannon et al., 2003).

All data corresponding to the GDSC dataset[1] (drug sensitivity, expression, copy number variations, single nucleotide variants, compounds, and cell lines) were obtained from the original website of the project as of September 2016. Drug sensitivity and transcriptomics data were used as provided. Genomics data were transformed into a binary matrix of genomic alterations vs. cell lines, where three different modifications for every gene were considered using the original source files: amplifications (ampSLCx), deletions (delSLCx), and variants (varSLCx). An amplification was annotated if there were more than two copies of at least one of the alleles for the gene of interest, and a deletion if at least one of the alleles was missing. Single nucleotide variants were filtered in order to exclude synonymous SNVs as well as nonsynonymous SNVs predicted not to be deleterious by either SIFT (Ng and Henikoff, 2001), Polyphen2 (Adzhubei et al., 2010), or FATHMM (Shihab et al., 2013).

LASSO Regression

LASSO regression analysis was performed using the "glmnet" R package (Friedman et al., 2010). Expression values for all genes in the dataset (17,419 genes in total) were used as input features. For each compound, the analysis was iterated 50 times over 10-fold cross validation. At each cross validation, features were ranked based on their frequency of appearance (number of times a feature has non zero coefficient for 100 default lambda possibilities). We then averaged the ranking across the 500 runs (50 iterations × 10 CV) in order to obtain a final list of genes associated to each compound. In this context, the most predictive gene for a certain drug does not necessarily have an average rank of one, even though its final rank is first.

[1] http://www.cancerrxgene.org/downloads

Elastic Net Regression

Elastic net regression analysis was performed using the "glmnet" R package (Friedman et al., 2010). Genomic data (copy number variations and single nucleotide variants) and transcriptional profiles of SLC and ABC genes across the cell line panel were used as input variables, either alone or in combination. Drug area under the curve (AUC) values were used as response. Elastic net parameters were fixed as follows: (i) alpha, the mixing parameter that defines the penalty, was set to 0.5 in order to apply an intermediate penalty between Ridge and LASSO, and (ii) lambda, the tuning parameter that controls the overall strength of the penalty, was determined individually for every model (drug) by optimizing the mean squared cross-validated error.

For each compound, 500 Elastic Net models were generated by a 100× fivefold cross-validation procedure. In order to assess model performance, the Concordance Index (CI) (Harrell et al., 1996; Papillon-Cavanagh et al., 2013) between the predicted and observed AUC values was calculated for each run, and then averaged across all models. This index estimates the fraction of cell line pairs for which the model correctly predicts which of the two is the most and least sensitive; hence CI values of 0.5 and 1 would indicate random and perfect predictors, respectively. Feature weights were calculated by normalizing the fitted model coefficients to the absolute maximum at every cross-validation run. The final list of features associated with each compound was built by computing the frequency of appearance of each feature in all the 500 models as well as its average weight. Features with positive weights are associated with a resistance phenotype to the compound, and negative weights are indicative of sensitivity.

RESULTS

SLC and ABCs as Drug Transporters

We collected data from public repositories as well as relevant publications to define the current knowledge on transport of chemical compounds by members of the SLC and ABC protein classes. A total of 493 compounds linked to 107 transporters were retrieved, which altogether formed a single large network with a few other smaller components (**Figure 1** and **Supplementary Table S1**).

Within the largest network and in agreement with previous reports (Nigam, 2015), three families are significantly enriched (hypergeometric test, FDR $\leq$ 0.05): the SLCO/SLC21 family of organic anion transporters (9/12 members) (Hagenbuch and Stieger, 2013), the SLC22 family of organic anion, cation, and zwitterion transporters (13/23) (Koepsell, 2013; Nigam, 2018), and the ABCC family of multidrug resistance transporters (8/13) (Vasiliou et al., 2009). Not surprisingly, ABCB1 (P-glycoprotein; MDR1), the very well-studied efflux pump known for its broad substrate specificity and mediation of resistance to a large amount of drugs (Aller et al., 2009), is the most connected transporter in the network, linked to more than 200 compounds. In particular, 106 compounds are connected exclusively with ABCB1 and 25 other are exclusively shared with ABCG2 (BCRP), another well-known transporter and the one with the second highest degree in the network (Robey et al., 2007; **Figure 1B**). Other top-connected SLCs include members of the above mentioned SLCO and SLC22 families, which also show several common substrates (e.g., SLCO1B1 and SLCO1B3 share 36 compounds, and SLC22A8 and SLC22A6 share 20), as well as members of the SLC15 family (SLC15A1 and SLC15A2, which share 22 compounds), involved in the transport of beta-lactam antibiotics and peptide-mimetics (Smith et al., 2013). In contrast to these cases, other transporters appear related to one or only a few compounds. One such case is SLC35F2, whose only reported substrate to date is the anti-cancer drug YM155 (sepantronium bromide) (Winter et al., 2014). Finally, while most chemical compounds appear linked to one or two transporters, a few others show higher connectivities (**Figure 1C**). A well-known example, methotrexate is transported by more than 20 different SLC and ABCs, including some belonging to families not commonly involved in drug transport, such as the folate carriers SLC19A1 and SLC46A1.

Transporter Expression Landscape in Cancer Cell Lines

The GDSC dataset contains expression data for 371 SLCs and 46 ABCs across a panel of ∼1,000 cell lines of different tissue origin. Each of these cell lines effectively express between 167 and 255 transporters, with a median value of 195 (**Figure 2A** and **Supplementary Table S2**). Although all together they cover almost the whole transporter repertoire (414/417), the distribution is clearly bimodal, with a common set of ∼130 transporters expressed in at least 900 cell lines and a more specific set of ∼140 expressed in less than 100 (**Figure 2B**). Among the most commonly expressed transporters, we find several members of the SLC25 (mitochondrial carriers) and SLC35 (nucleoside-sugars transporters) sub-families, the two largest among SLCs, as well as several members of the SLC39 family of zinc transporters. On the other end, many members of the SLC22 family, a large and well known group of proteins involved in the transport of drugs, as well as the SLC6 family, a well-studied family of neurotransmitter transporters, show a more specific expression pattern. As for ABCs, it is worth highlighting that subfamilies A and C present half of their members in the set of transporters of specific expression, while subfamily B has members in both sets.

When looking at actual expression values across the panel, some of the commonly expressed transporters coincide with those of highest expression (**Figure 2C**). The most extreme cases are SLC25A5, SLC25A3, SLC25A6, and SLC38A1, which present very similar maximum and median values across the cell line panel. On the contrary, other transporters such as SLC26A3, SLC17A3, or SLC38A11 present a much wider range of expression, being amongst the highest expressed in some cell lines but completely absent from others.

Finally, substantial differences become apparent when considering transporter expression patterns according to the tissue of origin of the GDSC cancer cell lines (**Figure 2D**). Cell lines belonging to the hematopoietic (blood) lineage, which includes leukemias, lymphomas, and myelomas, present

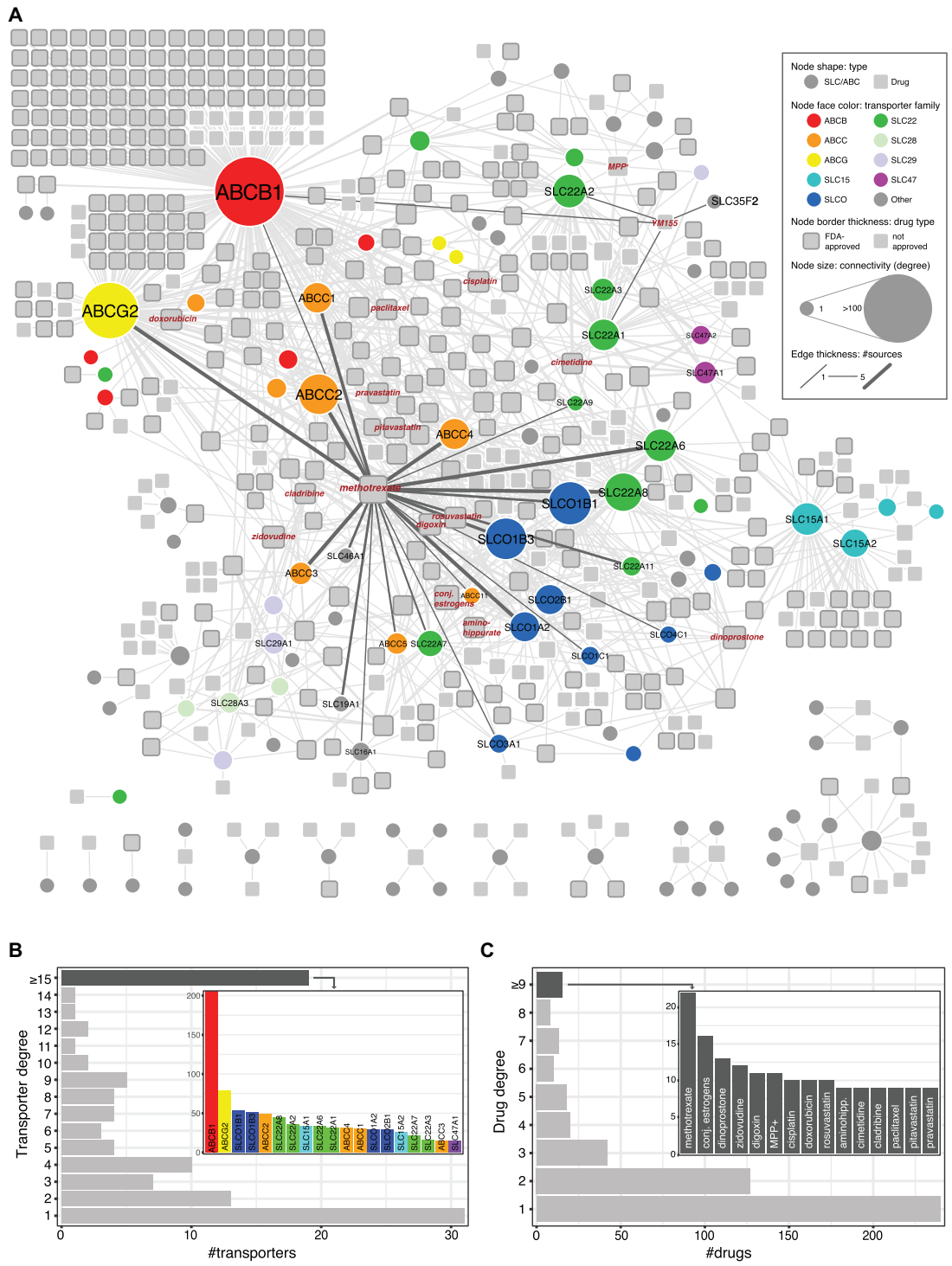


FIGURE 1 | (A) Network visualization of known SLC/ABC-mediated drug transport cases. *Circular* nodes represent SLC and ABC transporters, and *squares* represent chemical compounds. Drugs approved by the FDA (Food and Drug Administration) are displayed with *thicker gray* borders. Edges connect transporters to compounds and their thickness indicates the number of sources supporting each connection (see section "Materials and Methods"). Size indicates node degree (number of edges incident to the node). Relevant transporter families are color coded. **(B)** Transporter degree distribution. The inlet bar chart displays the transporters connected to at least 15 compounds. Bar colors correspond to transporter families in **(A)**. **(C)** Same as **(B)** for drugs.

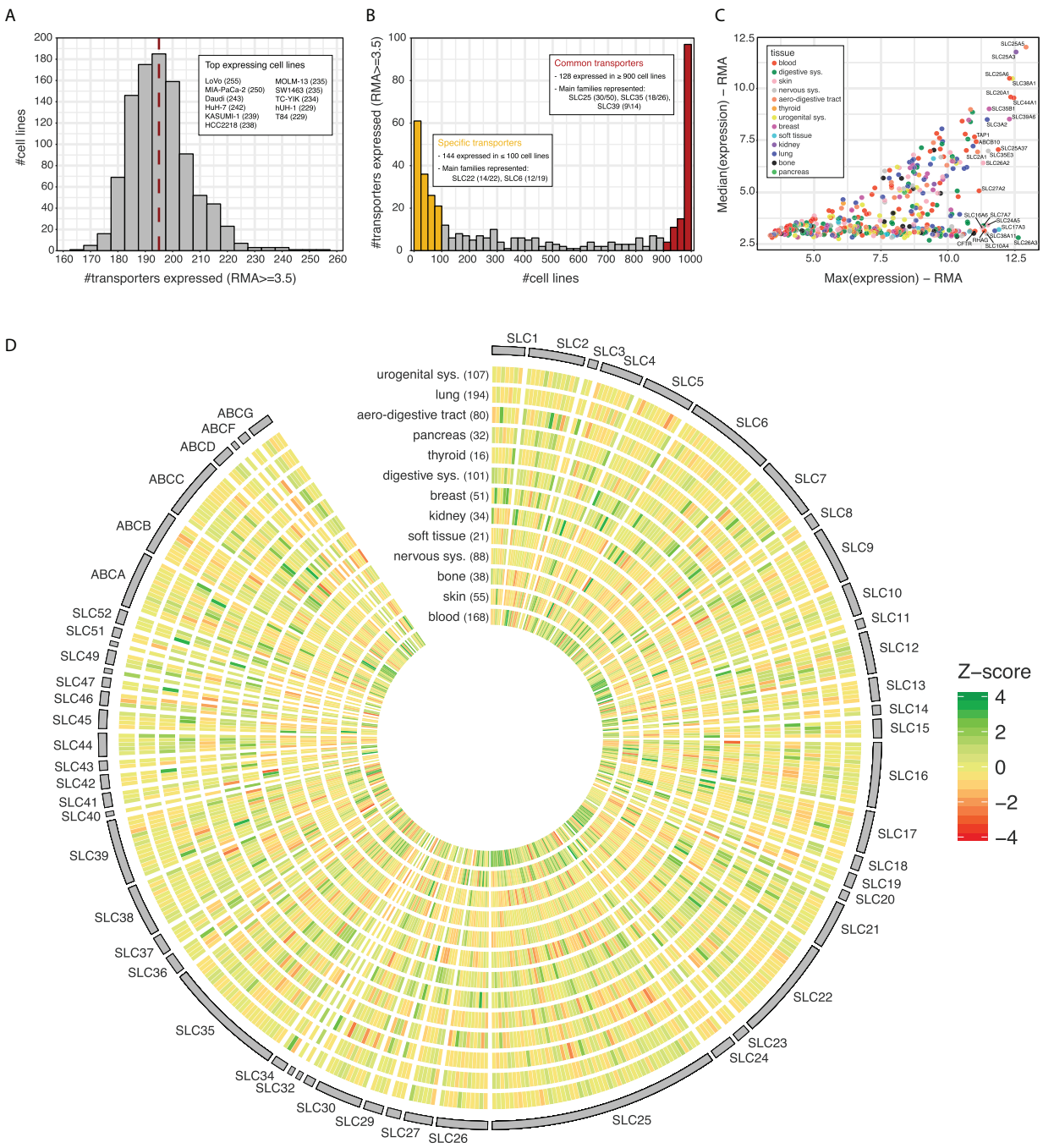


FIGURE 2 | (A) Number of transporters (SLCs and ABCs) expressed across cell lines in GDSC dataset. A cut-off of 3.5 in RMA units is set to consider a gene as expressed (~73% genes expressed). The *red line* indicates the median number of transporters expressed per cell line. The inlet lists the 11 cell lines expressing the highest number of transporters, indicated between parentheses. **(B)** Number of cell lines expressing each of the transporters. The color bars and inlets indicate sets of transporters showing more common or specific expression across the panel. **(C)** Median expression vs. maximum expression for each transporter across the cell line panel. *Color* indicates the tissue of origin of the cell line presenting the maximum expression for the transporter. **(D)** Transporter Z-scores of the average expression values within each tissue. Tissue names with number of cell lines between parenthesis are indicated on the x-axis. Transporters are ordered alphabetically by family and name.

the largest proportion of transporters with highest average expression values (28%), as indicated by Z-score, followed by cancer cell lines belonging to skin, kidney, and the digestive system. Interestingly, kidney cell lines also present the largest number of transporters with low expression values, pointing to a very wide range of expression and high specificity in those cells.

LASSO Regression Shows Importance of SLC Genes Across Whole Genome

We investigated the importance of SLC and ABC transporters for drug response by applying regularized linear regression on the GDSC dataset. To this end, we first built LASSO models of sensitivity to each compound based on genome-wide gene expression levels (17,419 genes in total) (Tibshirani, 1996), and then looked for cases where a transporter ranked as the top (first) predictor (see section "Materials and Methods"). The decision to focus exclusively on the top predictor is supported by a literature search. Indeed, the average number of PubMed publications containing both the drug and the gene name was over 40 in the case of top predictors, falling down to below 10 for the ones ranked second (**Supplementary Figure S1**).

Consistent with their well-characterized role as drug-transporters, the multi-drug resistance pump ABCB1, as well as ABCG2, were the main predictors of resistance to a large number of drugs (**Table 1**). More interestingly, several compounds had an SLC as their best predictor (**Table 2**). Among them, and in concordance with previous expression-sensitivity data (Rees et al., 2016), we find the sensitive association of sepantronium bromide (YM155) and SLC35F2, its main known importer (Winter et al., 2014). Another sensitive association involving SLC35F2 links this transporter to NSC-207895, a MDMX inhibitor (Wang et al., 2011). Dimethyloxalylglycin (DMOG), a synthetic analog of α-ketoglutarate that inhibits HIF prolyl hydroxylase (Zhdanov et al., 2015), showed association to two SLCs: monocarboxylate transporter SLC16A7 (MCT2) was the main predictor for sensitivity to this compound, while creatine transporter SLC6A8 (CT1) was associated with resistance. However, due to the high IC50 values of DMOG (in the millimolar range), this association is unlikely to be clinically relevant. Finally, cystine-glutamate transporter SLC7A11 (Blomen et al., 2015) is associated to resistance to the ROS-inducing drugs Shikonin, (5Z)-7-Oxozeaenol and Piperlongumine. This is in agreement with previous studies that highlighted a positive correlation of the expression of this transporter and resistance to several drugs via import of the cystine necessary for glutathione balance maintenance (Huang et al., 2005).

TABLE 1 | LASSO ABC-drug top associations.

LASSO top hits, all 17,419 genes used	Top sensitive associations (average rank)	Top resistant associations (average rank)
ABCB1		YM155 (1)
		Paclitaxel (1.1)
		BI-2536 (6.0)
		A-443654 (32)
		Vinorelbine (1)
		Thapsigargin (20)
		AT-7519 (1.8)
		WZ3105 (1)
		PHA-793887 (2.2)
		GSK690693 (15)
		Vinblastine (1.1)
		Docetaxel (1.2)
		ZM447439 (77)
		ZG-10 (1.3)
		QL-VIII-58 (1)
		QL-XII-61 (9.7)
ABCG2		CUDC-101 (12)
		THZ-2-102-1 (1.8)
ABCA10	STF-62247 (20)	
	FR-180204 (22)	

Elastic Net Regression Identifies Transporter–Drug Relationships

In order to further explore SLC and ABC involvement in drug response, we decided to build new predictive models using Elastic Net regression based on transporter molecular data only. Assessment of model performance was done by cross-validation using the CI (see section "Materials and Methods")

We considered different predictors to build the models: genomics (copy number variations and single nucleotide variants), transcriptomics (gene expression) and a combination of both. Among these, gene expression resulted to be most predictive, in agreement with previous reports (Aydin et al., 2014; **Figure 3A**). A total of 139 (53%) of the 265 drugs included in the dataset had predictive models with a CI higher than 0.60, and 36 (14%) higher than 0.65 (**Figure 3B**). For those drugs, we then ranked genes based on their frequency of appearance in the cross-validated models (indicative of the robustness of the association) and their average weight (indicative of the strength of the association as well as its direction). In this context, increased levels of transporter expression could therefore be associated with either sensitivity or resistance to the drug, for example, through its uptake or efflux, respectively (**Figure 3C**). Among the top ranked transporter–drug associations, we identified several known cases of drug transport. For instance, the strongest sensitivity association with sepantronium bromide (YM155) corresponded again to SLC35F2. Similarly, the strongest resistance association for this drug was ABCB1, which includes YM155 among its many substrates (Lamers et al., 2012; Voges et al., 2016; Radic-Sarikas et al., 2017). Another example was methotrexate, for which

TABLE 2 | LASSO SLC-drug top associations.

LASSO top hits, all 17,419 genes used	Top sensitive associations (average rank)	Top resistant associations (average rank)
SLC16A7	DMOG (1)	
SLC6A8		DMOG (40)
SLC30A2		CP724714 (28)
SLC35F2	YM155 (2.24)	
SLC35F2	NSC-207895 (9.5)	
SLC7A11		Shikonin (2)
SLC7A11		(5Z)-7-Oxozeaenol (12)
SLC7A11		Piperlongumine (12)

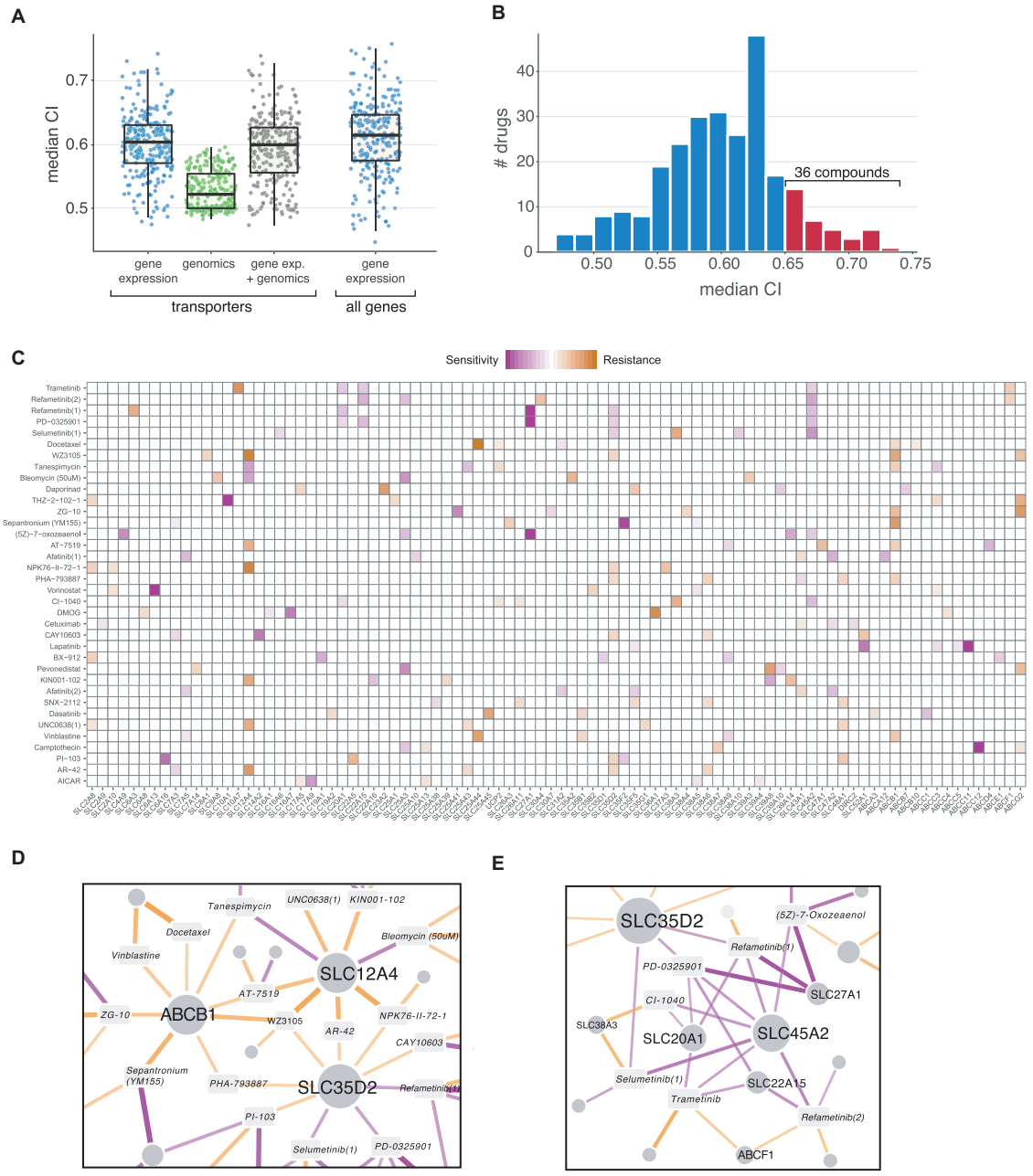


FIGURE 3 | (A) Comparison of Elastic Net regression performance (Concordance Index) using different input data: gene expression, genomics (CNVs and SNVs) and a combination of both. **(B)** CI value distribution using gene expression as input. *Red bars* indicate drugs with a median CI higher than 0.65, which were selected for subsequent analysis. **(C)** Elastic Net results for drugs with the highest CI values. The top five associations are shown for each compound. *Purple* indicates associations linked to sensitivity (higher expression value confers sensitivity to the compound), and *orange* indicates resistance. **(E)** Network representation of three transporters appearing as "hubs" (e.g., connected to several different compounds) in the results, including the well-known multidrug resistance protein ABCB1. **(D)** Same as **(E)** for MEK inhibitors, which show a similar association pattern.

the folate transporter SLC19A1, known to mediate its import (Zhao et al., 2011), ranked second for sensitivity association (**Supplementary Table S3**).

Two major patterns are apparent in the set of top-ranking associations: genes showing similar profiles of resistance or sensitivity across several different and unrelated compounds as well as groups of genes showing a similar profile in relation to a functionally related class of drugs (**Figure 3C**).

A prototypical case of the first pattern is ABCB1, which is associated with resistance phenotypes to several compounds (**Figure 3D**). Together with the aforementioned YM155, resistance relationships were predicted for known

substrates vinblastine and docetaxel (Fletcher et al., 2010), 17-AAG/Tanespimycin (Huang et al., 2007) and AT-7519 (Cihalova et al., 2015) as well as other not previously associated compounds such as ZG-10 (a JNK1 inhibitor), the CDK2/5/7 inhibitor PHA-793887 and the broad kinase inhibitor WZ3105. Similar to ABCB1, other transporters showed multiple resistance and sensitivity associations to different compounds, particularly kinases and chromatin-related enzymes. Two of these "hubs" were SLC12A4/KCC1, a potassium-chloride cotransporter involved in cell volume homeostasis (Arroyo et al., 2013), and SLC35D2, an activated sugar transporter localized in the Golgi (Song, 2013).

As an example of the second class of associations, some of the best models were achieved for compounds belonging to the MEK inhibitor drug class (Trametinib, Selumetinib, Refametinib, CI-1040, PD-0325901, (5Z)-7-oxozeaenol), which showed very similar patterns, with sensitivity associated to SLC45A2, SLC27A1, SLC20A1, and SLC22A15 (**Figure 3E**). SLC45A2 has been related to melanin synthesis and it is highly expressed in melanomas (Park et al., 2017), a cancer type sensitive to MEK inhibitors. Interestingly, SLC20A1/PiT1, a sodium-dependent phosphate transporter (Olah et al., 1994), was previously shown to regulate the ERK1/2 pathway independently of phosphate transport in skeletal cells (Bon et al., 2018). SLC27A1, a long-chain fatty acid transporter, and SLC22A15, an orphan member of the well-known family of cationic transporters involved in the transport of different compounds, were not previously associated with this drug class.

Finally, we also observed a strong sensitivity relationship between expression levels of the amino acid transporter SLC7A5/LAT1 and the Her2 and EGFR kinase inhibitors Afatinib, Gefitinib, and Bosutinib (**Figure 2C**), consistent with previously published data (Timpe et al., 2015).

DISCUSSION

Transporters of the ABC and SLC superfamilies are increasingly recognized as key players in determining the distribution and metabolism of drugs and other xenobiotic compounds as they possess distinct and extremely variable expression patterns across cell lines and tissues (O'Hagan et al., 2018). Moreover, they have been implicated in the development of resistance to several chemotherapeutic drugs (Fletcher et al., 2010). A survey of currently known drug transport relationships revealed that only a fifth of the more than 500 SLCs and ABCs have been described to be involved in the transport of drugs. These transporters appear to be very unevenly distributed, with some genes and families considerably more represented and better connected than others (**Figure 1**). This is the case for several members of the ABCB, ABCC, SLCO, and SL22 sub-families. Similarly, while compounds such as methotrexate are linked to more than 20 transporters, most drugs are connected to only one.

To further expand this network, we took advantage of the expression and drug sensitivity data available within the GDSC project. We started by characterizing the expression patterns of SLCs and ABCs in the GDSC panel of ∼1,000 cancer cell lines, covering thirteen different tissues of origin (**Figure 2**). Roughly 80% of SLCs and 90% of ABCs were included in the datasets and we observed a bimodal distribution of their expression, with similarly sized sets of transporters either present in most cell lines or specific to a few. In particular, a broad spectrum of expressed transporters was detected in cell lines derived from the hematopoietic (blood) lineage as well as in cell lines derived from skin, kidney, and the digestive system. A large variability in the level of expression was also observed within the superfamilies, consistent with what recently reported by another recent study (O'Hagan et al., 2018).

We then implemented a linear regression-based approach to identify the set of transporters associated with sensitivity to each compound across all cell lines. Previous reports undertook a similar approach to identify associations of the ABC (Szakacs et al., 2004) and SLCO/SLC22 (Okabe et al., 2008) families with drug sensitivity within a limited set of about 60 cell lines. We now extended these results to a much more comprehensive set of cell lines while implementing regularized linear regression approaches. In a first step, LASSO regression was used to assess genome-wide importance of transporters as predictors for drug sensitivity. The choice of the LASSO method was motivated by its ability to shrink a large number of coefficients to zero, ideal for models that make use of thousands of predictors. Moreover, being a linear regression method, it can account for both positive and negative interactions (i.e., resistance and sensitivity, for example, by export and import in the case of a transporter), while at the same time being more interpretable than more complex models. Subsequently, we based our analysis on transporter data only. By removing the effect of other genes in the models, we could prioritize compounds that show a stronger dependency on transporters, as well as to analyze potential cooperative interactions among them. We used in this case Elastic Net regression, a generalization of the LASSO that overcomes some of its limitations and that has already been applied in similar contexts (Zou and Hastie, 2005; Barretina et al., 2012; Iorio et al., 2016).

We identify a large set of drug–transporter associations roughly split between sensitivity and resistance relationships (**Figure 3** and **Tables 1**, **2**). Known associations involving, for example, ABCB1 expression levels with increasing resistance to several unrelated compounds as well as known interactions such as the associations between antifolates and SLC19A1 or YM155 and SLC35F2 were clearly identified. Interestingly, we also observed cases were, similarly to ABCB1, a single gene was associated with several compounds, possibly as a result of an alteration of the general metabolic state of the cell. We also observed the opposite scenario, with several genes associated with a functionally related class of compounds as in the case of the MEK inhibitors and the genes SLC45A2, SLC27A1, SLC20A1, and SLC22A15. To our knowledge, no transporter has so far been identified for any member of this class of compounds, and while the association with the skin-specific SLC45A2 transporter is likely the result of the high sensitivity of melanoma cell lines to these drugs, other associations are more difficult to interpret.

We propose the gene list reported here as a means of prioritizing transporters that could explain the transport and

pharmacodynamics of the associated compounds. While in many cases these associations could be due to indirect effects, such as a change in the metabolic state of the cells that renders them more sensitive or resistant to a compound, some might correspond to actual import or export processes. Further validation, for example, modulating the expression levels of the transporters or by transport assays, will be necessary in order to confirm and distinguish such different scenarios. Finally, the power of the analysis could also be increased by larger datasets, for instance including additional compounds, as well as by orthogonal or more accurate measurements. Availability of such pharmacogenomics datasets will be of critical importance for the further identification and characterization of transporter–drug associations.

In conclusion, we provide here an overview of the known ABC- and SLC-based drug transport relationships and expand this with an *in silico*-derived ranking of transporter–drug associations, identifying several novel and potential interesting cases. On the one hand, these new interactions offer new insights into the mode of drug transport across membranes, as well as provide initial structure activity relationships (SARs) for natural ligands, still unknown for many of these transporters. On the other hand, as many of the compounds involved in our analysis are clinically approved or candidates for oncological treatments, we hope that this study will provide novel hypotheses that could illuminate how transporters affect their pharmacodynamics and pharmacokinetics, as well as point to potential interactions with other compounds transported by the same proteins (e.g., in combination treatments), eventually leading to more specific and effective therapies.

AUTHOR CONTRIBUTIONS

AC-R and MY performed the data analysis. EG, MB, JS-R, and GS-F provided scientific insight and project supervision. AC-R, EG, MY, JS-R, and GS-F wrote the manuscript.

FUNDING

CeMM and the Superti-Furga laboratory were supported by the Austrian Academy of Sciences (GS-F and AC-R).

ACKNOWLEDGMENTS

We acknowledge receipt of third-party funds from the Austrian Science Fund (FWF I2192-B22 ERASE, AC-R and FWF P29250-B30 VITRA, EG), and from the JRC for Computational Biomedicine which was partially funded by Bayer AG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2018.01011/full#supplementary-material

FIGURE S1 | PubMed search of drug gene associations.

TABLE S1 | Known SLC/ABC-mediated drug transport cases.

TABLE S2 | Transporter expression landscape in cancer cell lines.

TABLE S3 | Elastic Net results.

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248

Alexander, S. P., Kelly, E., Marrion, N., Peters, J. A., Benson, H. E., Faccenda, E., et al. (2015). The concise guide to pharmacology 2015/16: transporters. *Br. J. Pharmacol.* 172, 6110–6202. doi: 10.1111/bph.13355

Aller, S. G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., et al. (2009). Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323, 1718–1722. doi: 10.1126/science.1168750

Arroyo, J. P., Kahle, K. T., and Gamba, G. (2013). The SLC12 family of electroneutral cation-coupled chloride cotransporters. *Mol. Aspects Med.* 34, 288–298. doi: 10.1016/j.mam.2012.05.002

Aydin, I., Weber, S., Snijder, B., Samperio Ventayol, P., Kuhbacher, A., Becker, M., et al. (2014). Large scale RNAi reveals the requirement of nuclear envelope breakdown for nuclear import of human papillomaviruses. *PLoS Pathog.* 10:e1004162. doi: 10.1371/journal.ppat.1004162

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi: 10.1038/nature1 1003

Blomen, V. A., Majek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096. doi: 10.1126/science.aac7557

Bon, N., Couasnay, G., Bourgine, A., Source, S., Beck-Cormier, S., Guicheux, J., et al. (2018). Phosphate (Pi)-regulated heterodimerization of the high-affinity sodium-dependent Pi transporters PiT1/Slc20a1 and PiT2/Slc20a2 underlies extracellular Pi sensing independently of Pi uptake. *J. Biol. Chem.* 293, 2102–2114. doi: 10.1074/jbc.M117.807339

Cesar-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., et al. (2015). A call for systematic research on solute carriers. *Cell* 162, 478–487. doi: 10.1016/j.cell.2015.07.022

Cihalova, D., Staud, F., and Ceckova, M. (2015). Interactions of cyclin-dependent kinase inhibitors AT-7519, flavopiridol and SNS-032 with ABCB1, ABCG2 and ABCC1 transporters and their potential to overcome multidrug resistance in vitro. *Cancer Chemother. Pharmacol.* 76, 105–116. doi: 10.1007/s00280-015-2772-1

Dobson, P. D., and Kell, D. B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat. Rev. Drug Discov.* 7, 205–220. doi: 10.1038/nrd2438

Fletcher, J. I., Haber, M., Henderson, M. J., and Norris, M. D. (2010). ABC transporters in cancer: more than just drug efflux pumps. *Nat. Rev. Cancer* 10, 147–156. doi: 10.1038/nrc2789

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01

Hagenbuch, B., and Stieger, B. (2013). The SLCO (former SLC21) superfamily of transporters. *Mol. Aspects Med.* 34, 396–412. doi: 10.1016/j.mam.2012.10.009

Harrell, F. E. Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 15, 361–387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Hediger, M. A., Clemencon, B., Burrier, R. E., and Bruford, E. A. (2013). The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol. Aspects Med.* 34, 95–107. doi: 10.1016/j.mam.2012.12.009

Hoglund, P. J., Nordstrom, K. J., Schioth, H. B., and Fredriksson, R. (2011). The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. *Mol. Biol. Evol.* 28, 1531–1541. doi: 10.1093/molbev/msq350

Huang, Y., Blower, P. E., Liu, R., Dai, Z., Pham, A. N., Moon, H., et al. (2007). Chemogenomic analysis identifies geldanamycins as substrates and inhibitors of ABCB1. *Pharm. Res.* 24, 1702–1712. doi: 10.1007/s11095-007-9300-x

Huang, Y., Dai, Z., Barbacioru, C., and Sadee, W. (2005). Cystine-glutamate transporter SLC7A11 in cancer chemosensitivity and chemoresistance. *Cancer Res.* 65, 7446–7454. doi: 10.1158/0008-5472.CAN-04-4267

Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754. doi: 10.1016/j.cell.2016.06.017

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951

Koepsell, H. (2013). The SLC22 family with transporters of organic cations, anions and zwitterions. *Mol. Aspects Med.* 34, 413–435. doi: 10.1016/j.mam.2012.10.010

Lamers, F., Schild, L., Koster, J., Versteeg, R., Caron, H. N., and Molenaar, J. J. (2012). Targeted BIRC5 silencing using YM155 causes cell death in neuroblastoma cells with low ABCB1 expression. *Eur. J. Cancer* 48, 763–771. doi: 10.1016/j.ejca.2011.10.012

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42, D1091–D1097. doi: 10.1093/nar/gkt1068

Morrissey, K. M., Wen, C. C., Johns, S. J., Zhang, L., Huang, S. M., and Giacomini, K. M. (2012). The UCSF-FDA transportal: a public drug transporter database. *Clin. Pharmacol. Ther.* 92, 545–546. doi: 10.1038/clpt.2012.44

Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res.* 11, 863–874. doi: 10.1101/gr.176601

Nigam, S. K. (2015). What do drug transporters really do? *Nat. Rev. Drug Discov.* 14, 29–44. doi: 10.1038/nrd4461

Nigam, S. K. (2018). The SLC22 transporter family: a paradigm for the impact of drug transporters on metabolic pathways, signaling, and disease. *Annu. Rev. Pharmacol. Toxicol.* 58, 663–687. doi: 10.1146/annurev-pharmtox-010617-052713

O'Hagan, S., Wright Muelas, M., Day, P. J., Lundberg, E., and Kell, D. B. (2018). GeneGini: assessment via the gini coefficient of reference "Housekeeping" genes and diverse human transporter expression profiles. *Cell Syst.* 6, 230.e1–244.e1. doi: 10.1016/j.cels.2018.01.003

Okabe, M., Szakacs, G., Reimers, M. A., Suzuki, T., Hall, M. D., Abe, T., et al. (2008). Profiling SLCO and SLC22 genes in the NCI-60 cancer cell lines to identify drug uptake transporters. *Mol. Cancer Ther.* 7, 3081–3091. doi: 10.1158/1535-7163.MCT-08-0539

Olah, Z., Lehel, C., Anderson, W. B., Eiden, M. V., and Wilson, C. A. (1994). The cellular receptor for gibbon ape leukemia virus is a novel high affinity sodium-dependent phosphate transporter. *J. Biol. Chem.* 269, 25426–25431.

Papillon-Cavanagh, S., De Jay, N., Hachem, N., Olsen, C., Bontempi, G., Aerts, H. J., et al. (2013). Comparison and validation of genomic predictors for anticancer drug sensitivity. *J. Am. Med. Inform. Assoc.* 20, 597–602. doi: 10.1136/amiajnl-2012-001442

Park, J., Talukder, A. H., Lim, S. A., Kim, K., Pan, K., Melendez, B., et al. (2017). SLC45A2: a melanoma antigen with high tumor selectivity and reduced potential for autoimmune toxicity. *Cancer Immunol. Res.* 5, 618–629. doi: 10.1158/2326-6066.CIR-17-0051

Radic-Sarikas, B., Halasz, M., Huber, K. V. M., Winter, G. E., Tsafou, K. P., Papamarkou, T., et al. (2017). Lapatinib potentiates cytotoxicity of YM155 in neuroblastoma via inhibition of the ABCB1 efflux transporter. *Sci. Rep.* 7:3091. doi: 10.1038/s41598-017-03129-6

Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. V., Gill, S., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* 12, 109–116. doi: 10.1038/nchembio.1986

Robey, R. W., Polgar, O., Deeken, J., To, K. W., and Bates, S. E. (2007). ABCG2: determining its relevance in clinical drug resistance. *Cancer Metastasis Rev.* 26, 39–57. doi: 10.1007/s10555-007-9042-6

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., et al. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden markov models. *Hum. Mutat.* 34, 57–65. doi: 10.1002/humu.22225

Smith, D. E., Clemencon, B., and Hediger, M. A. (2013). Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol. Aspects Med.* 34, 323–336. doi: 10.1016/j.mam.2012.11.003

Song, Z. (2013). Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Mol. Aspects Med.* 34, 590–600. doi: 10.1016/j.mam.2012.12.004

Sprowl, J. A., and Sparreboom, A. (2014). Uptake carriers and oncology drug safety. *Drug Metab. Dispos.* 42, 611–622. doi: 10.1124/dmd.113.055806

Sugano, K., Kansy, M., Artursson, P., Avdeef, A., Bendels, S., Di, L., et al. (2010). Coexistence of passive and carrier-mediated processes in drug transport. *Nat. Rev. Drug Discov.* 9, 597–614. doi: 10.1038/nrd3187

Szakacs, G., Annereau, J. P., Lababidi, S., Shankavaram, U., Arciello, A., Bussey, K. J., et al. (2004). Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. *Cancer Cell* 6, 129–137. doi: 10.1016/j.ccr.2004.06.026

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288.

Timpe, L. C., Li, D., Yen, T. Y., Wong, J., Yen, R., Macher, B. A., et al. (2015). Mining the breast cancer proteome for predictors of drug sensitivity. *J Proteomics Bioinform.* 8, 204–211. doi: 10.4172/jpb.1000370

Vasiliou, V., Vasiliou, K., and Nebert, D. W. (2009). Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics* 3, 281–290. doi: 10.1186/1479-7364-3-3-281

Voges, Y., Michaelis, M., Rothweiler, F., Schaller, T., Schneider, C., Politt, K., et al. (2016). Effects of YM155 on survivin levels and viability in neuroblastoma cells with acquired drug resistance. *Cell Death Dis.* 7:e2410.doi:10.1038/cddis.2016.257

Wang, H., Ma, X., Ren, S., Buolamwini, J. K., and Yan, C. (2011). A small-molecule inhibitor of MDMX activates p53 and induces apoptosis. *Mol. Cancer Ther.* 10, 69–79. doi: 10.1158/1535-7163.MCT-10-0581

Winter, G. E., Radic, B., Mayor-Ruiz, C., Blomen, V. A., Trefzer, C., Kandasamy, R. K., et al. (2014). The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat. Chem. Biol.* 10, 768–773. doi: 10.1038/nchembio.1590

Zhao, R., Diop-Bove, N., Visentin, M., and Goldman, I. D. (2011). Mechanisms of membrane transport of folates into cells and across epithelia. *Annu. Rev. Nutr.* 31, 177–201. doi: 10.1146/annurev-nutr-072610-145133

Zhao, Y., Butler, E. B., and Tan, M. (2013). Targeting cellular metabolism to improve cancer therapeutics. *Cell Death Dis.* 4:e532. doi: 10.1038/cddis.2013.60

Zhdanov, A. V., Okkelman, I. A., Collins, F. W., Melgar, S., and Papkovsky, D. B. (2015). A novel effect of DMOG on cell metabolism: direct inhibition of mitochondrial function precedes HIF target gene expression. *Biochim. Biophys. Acta* 1847, 1254–1266. doi: 10.1016/j.bbabio.2015.06.016

Zhou, F., Zhu, L., Wang, K., and Murray, M. (2017). Recent advance in the pharmacogenomics of human Solute Carrier Transporters (SLCs) in drug disposition. *Adv. Drug Deliv. Rev.* 116, 21–36. doi: 10.1016/j.addr.2016.06.004

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.5 Interlude II

Following the hypothesis that chemical compounds might need specific SLC transporters to gain access to cells, we set out to experimentally uncover new SLC-drug dependencies by means of a CRISPR-based genetic screen.

To this aim, we first selected a diverse set of 60 cytotoxic compounds comprising mostly approved drugs that distributed with no apparent bias across the complete drug chemical space. A CRISPR-Cas9 library was then constructed targeting 394 human SLC genes and used to infect haploid cells in order to generate a pool of single SLC knock-outs. Next, we treated the whole KO population with multiple concentrations of the selected compounds and interrogated samples for sgRNA abundance using next-generation sequencing. Significantly enriched sgRNAs in test samples versus controls were identified using a standard differential expression analysis approach followed by an aggregation method that led to a final list of gene candidates. Finally, validation of the most relevant candidates enabled us to identify a number of SLC-drug dependencies, which might consist in both direct drug uptake events as well as indirect associations in which SLCs modulate drug response by affecting cellular metabolism. The significant proportion of compounds for which we found SLC associations suggests a relevant and widespread role for SLC transporters in determining drug action and potency.

2.6 Manuscript #3

Girardi E., César-Razquin A, Papakostas K, Lindinger S, Konecka J, Hemmerich J, Kickinger S, Fiume G, Ringler A, Lardeau C-H, Kandasamy RK, Kubicek S, Ecker G, and Superti-Furga G () A Functional Survey of Solute Carrier-Drug Interactions. *Unpublished.*

Detailed author contribution: E.G. designed the project, performed experiments, analyzed data and wrote the manuscript; A. César-Razquin performed screen bioinformatic and statistical analysis, and contributed to figure design; K.P., S.L, J.K., G.F., A.R., C-H. L., and S.Ku. performed experiments and/or analyzed data; RK.K. designed negative control sgRNAs; J. H., S. Ki., and G.E. performed chemoinformatic analyses; G.S.F. supervised the study.

A functional survey of Solute Carrier-drug interactions

Enrico Girardi[1], Adrián César-Razquin[1], Konstantinos Papakostas[1], Sabrina Lindinger[1], Justyna Konecka[1], Jennifer Hemmerich[2], Stefanie Kickinger[2], Giuseppe Fiume[1], Anna Ringler[1,3], Charles-Hugues Lardeau[1,3], Richard Kumaran Kandasamy[1,4], Stefan Kubicek[1,3], Gerhard Ecker[2], Giulio Superti-Furga[1,5] *

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

2. Department of Pharmaceutical Chemistry, University of Vienna, Vienna, Austria

3. Christian Doppler Laboratory for Chemical Epigenetics and Antiinfectives, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria.

4. Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

5. Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria

* Correspondence: GSuperti@cemm.oeaw.ac.at

Abstract

The activity and potency of a drug is inherently affected by the metabolic state of its target cell. Solute Carriers (SLCs) represent the largest family of transmembrane transporters and constitute major determinants of cellular metabolism. Moreover, SLCs have been shown to be required for the uptake of an increasing list of compounds into cells. To identify novel transporter-drug associations, we performed a series of genetic screens in the haploid human cell line HAP1 using a set of 60 cytotoxic compounds representative of the chemical space populated by approved drugs. By using a SLC-focused CRISPR/Cas9 lentiviral library we identified several transporters whose absence induced resistance to the drug tested. Both known (e.g. antifolates and the folate transporter SLC19A1, nucleoside analogs and SLC29A1) as well as novel interactions (e.g. artemisinin derivatives and SLC11A2/SLC16A1, cisplatin and SLC35A2/SLC38A5) were identified for a significant proportion of the compounds screened, suggesting an important and widespread role for SLCs in determining cellular activity and uptake of cytotoxic drugs and providing an experimentally validated set of SLC-drug associations for a number of clinically relevant compounds.

Introduction

Drugs and drug-like compounds targeting intracellular proteins and processes are inherently influenced by the metabolic state of the cell they enter. Cellular metabolism influences the rates of drug uptake and extrusion through the presence of transmembrane transporters, the availability of cofactors and target(s) and the processing of prodrugs into active forms. Moreover, drug modifying enzymes (DMEs), such as members of the cytochrome C family and glucosyltransferases[1], add functional groups to xenobiotic compounds, facilitating their removal from the cell and eventually from the organism.

Most of what is known about the uptake of drugs by membrane-bound transporters stems from the analysis of drug disposition in the kidney, liver, intestine and blood-brain-barrier, with a particular focus on the entry and exit of pharmacological agents from the blood circulation[2,3]. In particular, two families of transporters have been previously shown to directly interact with drugs: ATP-binding cassette transporters (ABCs)[4] and Solute Carriers (SLCs)[5]. ABC transporters are generally involved with the extrusion of drugs, while SLCs have been mostly described to be involved in compound uptake, although exceptions to this rule exist such as in the case of MATE transporters[6]. Notably, SLCs represent a largely understudied family, counting more than 400 members of which ~30% are still considered orphan[3]. SLCs are divided on subfamilies based on sequence similarity and have been shown to transport a variety of molecules, ranging from nucleotides, sugars and lipids to aminoacids and peptides[7,8], often with overlapping specificities. Consistent with their critical role in drug absorption and excretion, considerable knowledge has accumulated on a few large subfamilies of SLCs prevalently expressed in kidney, liver and intestine, such as the SLC22 and SLCO families[9,10], and there is ample consensus in ascribing an important role for these transporters in pharmacokinetics of several drugs, which has been corroborated by several pharmacogenomic polyphormisms[11]. Accordingly, FDA and EMA now recommend testing of several ABC and SLC22/SLCO members for clinical drug interaction studies[12]. However, how

large the implication of membrane-bound transporters may be in the uptake and disposition of drugs at the target cell level, such as in muscle, brain or tumor cells, remains a matter of debate[13–15]. Several drugs have been reported to depend on protein carriers to enter cells, with prominent cases such as the family of antifolate drugs (e.g. methotrexate, pralatrexate, raltitrexed) and the folate transporters SLC19A1/RFC1 (reduced folate carrier 1) and SLC46A1/PCFT (proton-coupled folate transporter)[16], or the nucleoside transporter SLC29A1/ENT1 (equilibrative nucleoside transporter 1) and several nucleoside analogs such as clofarabin, gemcitamine and fluorouracil[17]. In parallel, modulation of transporter activity or expression levels has been shown to affect the efficacy of drugs, independently of direct uptake events, thorough their effects on cellular metabolic processes such as glycolysis and oxidative phosphorylation[18,19].

Genetic screening offers a powerful tool to identify both direct and indirect interactions between a gene and a specific phenotype. Using insertional mutagenesis, we recently demonstrated that the presence of the intact SLC35F2 gene was the major determinant of the uptake of sepantronium bromide (YM155), a small molecule displaying anti-tumor activity in vitro and in vivo, in a variety of cell lines[20]. We hypothesized that, given the lack of information and molecular tools available for the solute carrier family[3], it could be possible to identify several cases of dependencies beyond the above-mentioned examples, as well as several cases in which SLCs affect indirectly the potency of drugs. Using a SLC-specific CRISPR/Cas9 KO library, an unbiased genetic screening of a chemically diverse set of 60 cytotoxic drugs allowed us to identify and validate a large number of SLC-compound associations, providing insights into both direct uptake events and indirect associations affecting the metabolism and mechanism of action of the drugs tested.

Results

Generation of a SLC-specific CRISPR/Cas9 library

In order to interrogate all SLC genes in an unbiased manner, we constructed a CRISPR/Cas9 library targeting 394 human SLC genes and pseudogenes with multiple sgRNAs per gene. Particular care was taken to avoid sgRNA with sequences sharing similarity with other SLC or ABC transporters. A set of negative control sgRNAs (predicted not to target any sequence in the genome) as well as a set of genes scoring as essential in the HAP1 and KBM7 cell lines based on previous insertional mutagenesis data[21] were also included in the pool (Fig 1a, Suppl Table 1). The resulting library consisted of 2609 unique sgRNAs, allowing for highly scalable and multiplexable screening and sequencing protocols. Presence of all sgRNAs was confirmed by NGS sequencing (Suppl Fig 1a). Comparison of plasmid samples with samples taken 9 days post-infection showed significant depletion of sgRNAs targeting the set of essential genes (54/120, p-value = 8.2×10^{-26}, Fisher's exact test, Suppl Fig 1b). No significant depletion or enrichment was observed for the set of negative control sgRNAs (21/120, enrichment p-value = 0.29, depletion p-value = 1, Fisher's exact test, Suppl Fig 1c). At the gene level, we identified several SLCs important for optimal fitness of HAP1 cells, including SLC35B1, the recently deorphanized ATP/ADP exchanger in the endoplasmic reticulum[22] and MTCH2, a mitochondrial carrier involved in the regulation of apoptosis[23] (Suppl Fig 1d). To validate the efficiency and specificity of our library, we screened for SLC genes responsible for resistance to YM155. Screening in HAP1 cells with 200nM YM155 for 72h resulted in a clear enrichment in sgRNAs targeting the SLC35F2 gene (Suppl Fig 1e), confirming that SLC35F2 is the sole SLC responsible for YM155 resistance and consistent with our previous results derived from gene-trapping experiments[20].

The SLC repertoire of HAP1 cells

Immortalized human cell lines typically express 150-250 SLC genes, with abundancy patterns resembling those of tissues[24,25]. For our studies, we chose HAP1 cells, a human cell line bearing considerable technical advantages[26,27]. HAP1 cells express 207 human SLC genes, as assessed by transcriptional profiling using RNA-Seq[28] (Fig 1b). Interestingly, these cells do

not express most members of the organic ion transporter SLC22 family that have been implicated in the uptake of drugs in kidney, gut and liver[29]. Moreover, by being haploid, loss-of-function phenotypes induced by CRISPR/Cas9 technology should be more easily interpretable, as they do not represent composite mutants of different alleles.

Identification of a set of cytotoxic drugs

For our genetic screens, we aimed at selecting a set of compounds representative of the chemical space populated by drugs. We therefore tested cytotoxicity of a set of 1812 compounds (2k library) including the CLOUD library[30] and the NIH Clinical Collection as well as sets of epigenetic modifiers and toxic compounds. A subset of 270 (14.9%) compounds was found to be cytotoxic in HAP1 cells at the tested concentration (toxic set) (Fig 1c). A 8-point dose-response curve was subsequently performed for each drug to determine IC_{50} values. A subset of 60 drugs, which also included additional 9 compounds involved in DNA-damage-based cytotoxicity, was then selected for screening with the CRISPR/Cas9 library (screen set). This set includes mostly approved drugs covering several different classes (Suppl Table 2).

Chemical space analysis

In order to assess if the selected set of compounds (screen set) was representative of the general drug-like space, a cheminformatics analysis was performed. Drugbank 5.1.1[31] was used as a reference of the known drug chemical space, and compared to all three above-mentioned compound sets (2k library, toxic and screen, Suppl Fig 1f). All sets were curated according to the same protocol (see methods) and 22 physicochemical 2D descriptors were calculated for every compound (Suppl Table 3). Comparison of individual descriptor mean and median values showed no significant bias across the four compound sets. In order to visualize the distribution of all compounds in the chemical space, a principal component (PC) analysis was performed. The first and the second PCs were able to explain 62.1 % of the variance of the data (Fig 1d). Descriptors contributing the most to the variance of PC1-2 were number of

heavy atoms, molecular weight, labute surface area, number of heteroatoms, number of saturated rings, number of H-bond donors and polar surface area (TPSA) (Suppl Fig 1g). Importantly, compounds of all sets were similarly distributed along the two first PCs. Overall, this analysis confirmed that there is no significant difference in the distribution of physicochemical properties of the compound sets used in this study, and that the final screen set can be considered representative of the general drug chemical space.

Genetic screening

We then infected haploid HAP1 cells with the SLC KO library to generate a pool of cells each lacking, on average, a specific SLC. The population was treated with multiple concentrations (generally one, three and ten times the measured IC_{50}) of the cytotoxic compounds for 72h. As expected, we retrieved all the samples treated with the IC_{50} concentrations, as well as 36/60 (60%) of the treatments at 3X IC_{50} and 22/60 (37%) of the 10X IC_{50} treatments. Enrichment was first calculated at the sgRNA level using DESeq2[32] (Fig 2a) and then aggregated at the gene level using a slightly modified version of the GSEA algorithm[33] (Fig 2b). When enrichment for SLC genes was calculated, we identified 201 SLC/drug associations involving 47 drugs (76 treatments) and 101 SLCs (Figure 2c, Suppl Fig 2a) at FDR <1%. Enrichment was more prominent for higher selective pressures than for IC_{50} treatment in terms of hit strength and most of the SLCs identified are expressed in HAP1 cells (93/101, 92%, Suppl Fig 2b).

Several clusters involving drugs belonging to the same classes were identified. One example is represented by the cluster of the antifolate drugs methotrexate, raltitrexed and pralatrexate, which all induced a strong enrichment in KOs of the reduced folate carrier RFC/SLC19A1. This transporter has been previously recognized to represent the main uptake route of these antimetabolites into cells[16]. In particular, pralatrexate was developed to exploit this entry route[34] and it shows exclusive enrichment for SLC19A1 in our screen (Fig 2c). Interestingly, within this cluster we find the structurally unrelated drug pentamidine, which is used for the treatment of African trypanosomiasis and leishmaniasis, as well as for the prevention and treatment of

pneumocystis pneumonia (PCP) in immunocompromised patients. The mechanism of action (MoA) of this drug is poorly understood but earlier reports suggested it might be involved with inhibition of the parasite dihydrofolate reductase[35]. Another cluster included the nucleoside-like drugs decitabine, cytarabine, 5-azacytidine and gemcitabine, which all showed enrichment for the nucleoside transporter ENT1/SLC29A1 (Fig 2c). SLC29A1 was previously reported to act as an importer of these compounds[17,36,37]. As HAP1 cells express very low levels/do not express the additional nucleoside transporters SLC29A2, SLC28A1 and SLC28A3 (Fig1b), loss of SLC29A1 results in an impaired uptake of these compounds within the cell. Interestingly for some (i.e. cytarabine and decitabine), but not all of these compounds, we detected enrichment of the mitochondrial phosphate transporter SLC25A3.

We also observed more specific interactions, such as the one between the antineoplastic drug mitoxantrone and the two transporters MATE1/SLC47A1 and MATE2/SLC47A2 (Fig 2c). Mitoxantrone was previously reported to inhibit the uptake/efflux of MATE1/SLC47A1 substrates[38,39], suggesting a direct interaction of this compound with these transporters.

Moreover, we observed several cases of interactions providing insights in the MoA or metabolic impact of a drug treatment. One case is the one of the artemisinin-derivatives artesunate and dihydroartemisinin, which showed an enrichment for the SLC11A2 and SLC16A1 genes (Fig 2c). These compounds, generally used for the treatment of malarial infections, have recently found use as antineoplastic agents[40,41]. Although the MoA is not fully understood, their cytotoxicity appears to be dependent on an iron/heme-dependent activation step and subsequent generation of ROS species[42]. SLC11A2, also known as DMT1 (divalent metal transporter 1), is an endosomal iron transporter which has been shown to control the pool of cytoplasmic iron[43]. SLC16A1, also known as MCT1 (monocarboxylate transporter 1) is a major lactate exporter that plays an important role in glycolytic metabolism[44] and could be directly involved in drug uptake or affect the ROS response to these compounds. Finally, we also observed a very strong enrichment of the transporters SLC35B2, a nucleoside-sugar Golgi

transporter[45], and SLC38A5[46], an amino acid transporter, upon treatment with the DNA-damaging agent cisplatin (Fig 2c).

Validation of selected SLC/drug associations by Multicolor Competition Assay

We selected a set of 34 SLC/drug interactions for experimental validation (Suppl Table 4), involving 21 drugs and 13 SLCs. In order to validate the interactions, we applied a FACS-based Multicolor Competition Assay (MCA), an approach that has been previously used to validate several forward genetics screen results[47]. In this assay, HAP1 cells carrying a sgRNA targeting a given SLC and an eGFP expression construct were mixed at 1:1 ratio with cells carrying an unspecific sgRNA (targeting the *Renilla spp luciferase* gene) and a mCherry construct. The mixed population was then treated with either vehicle or the cytotoxic compound at 1-3 times the IC_{50} and the ratio of GFP+/mCherry+ determined by FACS three and ten days later (Fig 3a). Two independent sgRNA were used for each gene targeted in order to avoid sgRNA-specific effects. Overall, this approach enabled us to validate several of the strongest interactions derived from the genetic screen (Fig 3b). In particular, we confirmed the strong effects of SLC19A1 and SLC29A1 loss on the resistance to antifolate and nucleoside analogs at both early (3 days) and late (10 days) timepoints. In addition, we also validated the effect of the loss of SLC20A1, a phosphate transporter, on methotrexate resistance and the effect of SLC25A3, a mitochondrial phosphate transporter, on cytarabine. We also observed strong and time-dependent enrichments of cells lacking SLC11A2 or SLC16A1 upon treatment with artesunate and dihydroartemisinin, as well as in the case of panobinostat and the amino acid transporter SLC1A5. Finally, we observed strong enrichment of SLC35B2- and SLC38A5-lacking cells upon cisplatin treatment. This effect was already discernible for SLC38A5 after three days of treatment and became clear for both genes after ten days. Overall, we were able to confirm the majority (19/34) of the associations tested, therefore validating the approach and results of the genetic screen.

Discussion

Transmembrane transporters represent a major class of metabolic genes involved in several cellular processes affecting drug potency and activity, including the uptake and extrusion of these xenobiotic compounds[2,29]. The identification of the orphan transporters SLC35F2 and SLC37A3 as the main mediators of the uptake of the cytotoxic compound YM155[20] and of the release of bisphosphonates from intracellular vesicles[48], respectively, provided clear examples of the use of genetic approaches to identify such relationships. However, despite the unambiguous involvement of SLC35F2 and SLC37A3 in drug transport, their substrate drugs are charged at physiological pH. It is therefore possible that these relationships are exceptional. Alternatively, most drugs thought to act on an intracellular target indeed require a membrane-bound transporter to gain access to the inside of cells. According to this hypothesis, the main reason of why it may so far not been common to identify obligate drug-transporter dependencies stems from the functional redundancy among transporters, particularly among the more than 400 SLC-encoding human genes, and the lack of an experimental set-up allowing for clear loss-of-function states[5,48]. We therefore reasoned that a focused forward genetic approach (Fig 1a, Suppl Table 1) could allow us to systematically survey how often SLCs affect compound activity and to identify new potential cases of interaction between a transporter and a drug or drug-like compound.

As read-out compatible with genetic screening, we opted for viability, as it is results in strong selective pressures and allows us to focus on cytotoxic or cytostatic compounds of major clinical relevance. We selected a set of 60 compounds including mainly approved drugs (Fig 1c, Suppl Table 2). A cheminformatics analysis showed no bias in physicochemical properties of this set compared to the DrugBank database, thus supporting its use as a set representative of the chemical space occupied by drugs (Fig 1d).

Antimetabolites such as antifolates and nucleoside-analogs scored strongly in our setting, recapitulating the known cases of drug uptake mediated by transporters such as SLC19A1 and

SLC29A1 (Fig 2c, 3b). Interestingly, we also identify several additional strong interactions across different drug classes, such as the role of the iron transporter SLC11A2 in determining resistance to artemisinin derivatives (Fig 2c, 3b). This is consistent with the correlation between intracellular iron levels and drug cytotoxicity previously suggested for these compounds[50]. We also validated interactions between artesunate/dihydroartemisinin and the monocarboxylate transporter SLC16A1 as well as between cisplatin and the transporters SLC35A2 and SLC38A5. The latter is particularly interesting as SLC38A5 is a glutamine transporter expressed at high levels in cells of hematopoietic origin and several studies reported that cisplatin-resistant cells are dependent on glutaminolysis[51,52]. Further studies are ongoing to determine how loss of these transporters results in increased resistance to cytotoxicity. Moreover, we observed several interactions involving key, often essential, transporters involved in major energetic pathways such as SLC2A1/GLUT1, the major glucose transporter at the plasma membrane, SLC25A3, the mitochondrial phosphate transporter, or MTCH2, a mitochondrial carrier involved in apoptosis regulation (Fig 3b). It has been shown that resistance to cytotoxic drugs often requires major metabolic rearrangements: e.g. glutaminolysis and cisplatin resistance[51,52], a switch to oxidative phosphorylation in cytarabine resistance[53] or drug-specific dependence on glycolysis[18]. The fact that several of the SLC-drug associations identified involve SLCs important for cellular fitness therefore speaks to the enormous metabolic pressure a cytotoxic drug imposes on a target cell.

In conclusion, we surveyed the role of SLCs on the cytotoxic activity of a diverse set of compounds, including several clinically approved drugs. We found significantly enriched SLCs for 80% of the molecules tested, pointing to a clear role of this family in determining cellular responses to a broad range of molecular insults. Importantly, we identified and validated several novel SLC-drug associations involving widely used drugs, such as cisplatin and artemisinin derivatives, which warrant further investigation and could provide important insights into transport events, MoAs and mechanism of resistance to these classes of compounds.

Materials and methods

Generation of a SLC-wide CRISPR/Cas9 lentiviral library

A set of single guide RNAs (sgRNAs) targeting 388 human SLC genes, generally with six sgRNA per gene, were manually selected (or generated) to include sequences with predicted high efficiency and specificity, as assessed in Doench *et al*[54], and to minimize targeting of other SLCs or of ABC transporters (Suppl Table 1). sgRNAs targeting six SLC pseudogenes (SLC7A5P1, SLC7A5P2, SLC9A7P1, SLC2A3P1, SLC25AP5, SLC35E1P1) for which transcription was previously reported in at least two expression datasets (FANTOM5, CCLE, ENCODE, Cosmic, GENCODE, Uhlen et al, Illumina)[55–61] were also included. An additional set of 120 sgRNAs targeting 20 genes essential in both KBM7 and HAP1 cells[21] based on the number of retroviral insertions observed were also selected (Suppl Table 1). Finally, a set of 120 non-targeting sgRNAs was included by generating random 20-mers and selecting for sequences with at least three (for the strong PAM NGG) or two (for the PAM NAG) mismatches from any genomic sequence with E-CRISP Evaluation[62]. Adapter sequences were added to the 5' and 3' sequences (5'prefix: TGGAAAGGACGAAACACCG, 3'suffix: GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC) to allow cloning by Gibson assembly in the lentiCRISPRv2 vector (Addgene #52961). The oligos were synthetized as a pool by LC Sciences (Houston, Texas, US). Full-length oligonucleotides (74 nt) were amplified by PCR using Phusion HS Flex (NEB) and size-selected using a 2% agarose gel (Primers: SLC_ArrayF TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGAC GAAACACCG, SLC_ArrayR ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCT AGCTCTAAAAC)

The vector was digested with BsmBI (NEB) for 1h @ 55˚C, heat inactivated for 20' at 80˚C, following by incubation with Antarctic phosphatase (NEB) for 30' at 37˚C. A 10 µl Gibson ligation reaction (NEB) was performed using 5 ng of the gel-purified inserts and 12.5 ng of the

vector, incubated for 1h at 50˚C and dialyzed against water for 30' at RT. The reaction was then transformed in Lucigen Endura cells and plated on two 245 mm plates. Colonies (equivalent to approximately 200X coverage) were grown at 32˚C for 16-20h hours and then scraped from the plates. The plasmid was purified with the Endo-Free Mega prep kit (Qiagen).

Library NGS sequencing

Initial amplification of the library for NGS sequencing was performed by a two-step PCR protocol as described in Sanjana *et al* [63]. Due to the presence of unspecific bands affecting the quality of the sequencing experiments, later samples were processed with a single-step PCR derived from Konermann *et al* [64]. The PCR primers used to add barcodes and Illumina adapters were modified to allow for double indexing of samples.

Enrichment analysis

sgRNA sequences were extracted from NGS reads, matched against the original sgRNA library index and counted using an in-house python script. A two-step approach was implemented in order to obtain a final list of enriched candidate genes. First, differential abundance of individual sgRNAs was estimated using DESeq2[32]. Models accounted for both treatment and time variables when time 0 samples were available; otherwise only the treatment factor was considered. Contrasts were performed individually for each treatment and dose vs controls (DMSO and untreated), and significance was tested using either one- or two-tailed Wald tests (i.e. alternative hypothesis LFC>0 for enrichment, and abs(LFC)>0 for enrichment or depletion, respectively). Then, sgRNAs were sorted by log2 fold change and aggregated into genes using Gene Set Enrichment Analysis (fgsea R package)[33,65]. To avoid false positives, only significant sgRNAs (p-value ≤ 0.05) were considered for enrichment, requiring also a minimum of two sgRNAs per gene. Gene enrichment significance was estimated by a permutation test using 10^8 permutations, and p-values were corrected for multiple testing using the Benjamini-Hochberg procedure (FDR).

Cell lines

HAP1 cells (Horizon Genomics) were grown in IMDM media (Gibco) supplemented with 10% FCS (Gibco) and 1% penicillin/streptomycin. For screening purposes, haploid cells were selected by FACS sorting after staining with Vybrant DyeCycle Ruby stain (Thermo Fisher Scientific), expanded for 3-5 days and frozen until further use. For CRISPR-based knockout cell lines, sgRNAs were designed using CHOPCHOP[63] and cloned into pLentiCRISPRv2 (Addgene, #52961), LGPIG (pLentiGuide-PuroR-IRES-GFP) or LGPIC (pLentiGuide-PuroR-IRES-mCherry)[47]. sgRen targeting Renilla luciferase cDNA was used as negative control sgRNA[47].

Drug cytotoxicity screens

To mimic the genetic screen conditions, HAP1 cells were infected with a lentiCRISPRv2 vector carrying a sgRNA targeting the *Renilla* luciferase gene and selected with puromycin selection (1μg/ml) for 7 days. WT and lenti-infected cells were screened against a library composed of 1812 compounds at a single concentration in the range of 10-50μM. Viability was measured by CellTiterGlo assay (Promega) after 72h of treatment. DMSO (0.1%) and Digitoxin (10 μM) were used a negative and positive controls to calculate cytotoxicity. Hits were defined as compounds giving more than 50% inhibition compared to DMSO controls. 8-point dose-response curves were performed to determine the IC_{50} values of the cytotoxic compounds in lentivirus-infected HAP1 cells.

Chemical space analysis

Data curation was performed using a KNIME 3.6.0[66] workflow which incorporates the python packages RDKit 2018.09.01[67] and MolVS 0.1.1[68] for handling and standardizing molecules (python 3.6.6[69] was used). First, all compounds were neutralized by adding or removing protons. Then, compounds were cleaned by standardizing the representation of all aromatic rings, double bonds, hydrogens, tautomers and mesomers. Thereafter, all salts and mixtures

were removed. In order to remove duplicates InChIKeys were calculated and all compounds were aggregated according to these InChIKeys. Chiral centers were also removed, as this stereochemistry information is often incorrectly assigned, which can lead to a lower detection rate of duplicates. Furthermore, only 2D descriptors were calculated, which cannot differentiate between enantiomers or diastereomers. All 22 descriptors were computed with the RDKit nodes available in KNIME 3.6.0. Data visualization was performed in Rstudio 1.1.463[70] with R 3.4.4[71]. Bar plots and violin plots were computed with ggplot2 3.1.0[72], the correlation of descriptors plot (Suppl Fig 1g) was computed with corrplot 0.84[73]. Principal component analysis (PCA) was performed with the R packages factoextra[74] and FactoMineR[75].

Genetic screens

Viral particles were generated by transient transfection of low passage, subconfluent HEK293T cells with the SLC-targeting library and packaging plasmids psPAX2, pMD2.G using PolyFect (Qiagen). After 24h the media was changed to fresh IMDM media supplemented with 10% FCS and antibiotics. The viral supernatant was collected after 48h, filtered and stored at -80C until further use. The supernatant dilution necessary to infect haploid HAP1 cells at a MOI (multiplicity of infection) of 0.2-0,3 was determined by puromycin survival after transduction as described in Sanjana *et al*[76]. HAP1 cells were infected in duplicates with the SLC KO library at high coverage (1000x) and after selection for 7 days with puromycin (1 μg/ml) an initial sample was collected to control for library composition. Cells were then treated with multiple concentrations (1X, 3X or 10X the IC_{50}) of the cytotoxic compounds or vehicle (DMSO or DMF) controls for 72h and, when surviving cells were present, cell samples collected from both treated and control samples.

Multicolor competition assay

Flow cytometry-based multi-color competition assays (MCA) were performed as described previously[47]. Briefly, HAP1 cells expressing LGPIC-sgRen (mCherry-positive) were mixed in 1:1 ratio with LGPIG (eGFP-positive) reporter cells containing sgRNAs targeting the gene of

interest. The mixed cell populations were incubated with vehicle or drug for up to 10 days. The respective percentage of viable (FSC/SSC) mCherry-positive and eGFP-positive cells at the indicated time points was quantified by flow cytometry. Samples were analysed on an LSR Fortessa (BD Biosciences) and data analysis was performed using FlowJo software (Tree Star Inc., USA).

Author contributions

E.G., G. S-F. conceived and designed the study. E.G., K.P., S.L., J.K., G.F., A.R., and C-H. L. performed experiments and analyzed data. A.C-R analyzed screening and validation data. J.H, S.K. and G.E. performed the chemoinformatic analysis. R.K.K. contributed to library design. S.K., G.E. and G.S-F. provided supervision.

Acknowledgments

We thank all members of the Superti-Furga laboratory for discussions and feedback. We are also grateful to the Biomedical Sequencing facility for advice on Illumina sequencing and to the Flow Cytometry Core Facility of the Vienna Medical University for help with FACS sorting. We further thank Dr. Bojan Villagos for graphical input and advice. We acknowledge support by the Austrian Academy of Sciences, ERC grant to G.S.-F. (ERC AdG 695214 GameofGates, E.G.), Austrian Science Fund (FWF I2192-B22 ERASE, A. C-R, FWF P29250-B30 VITRA, E.G, J.K, G.F) and a Marie Sklodowska-Curie fellowship to E.G. (MSCA-IF-2014-661491).

References

1. Testa, B. & Krämer, S. D. The Biochemistry of Drug Metabolism – An Introduction: Part 1. Principles and Overview. *Chem. Biodivers.* **3**, 1053–1101 (2006).
2. Nigam, S. K. What do drug transporters really do? *Nat. Rev. Drug Discov.* (2014). doi:10.1038/nrd4461
3. César-Razquin, A. *et al.* A Call for Systematic Research on Solute Carriers. *Cell* **162**, 478–487 (2015).
4. Fletcher, J. I., Haber, M., Henderson, M. J. & Norris, M. D. ABC transporters in cancer: more than just drug efflux pumps. *Nat. Rev. Cancer* **10**, 147–156 (2010).
5. Hediger, M. A., Clémençon, B., Burrier, R. E. & Bruford, E. A. The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Mol. Aspects Med.* **34**, 95–107 (2013).
6. Motohashi, H. & Inui, K. Multidrug and toxin extrusion family SLC47: Physiological, pharmacokinetic and toxicokinetic importance of MATE1 and MATE2-K. *Mol. Aspects Med.* **34**, 661–668 (2013).
7. Hediger, M. A. *et al.* The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pfl□gers Arch. Eur. J. Physiol.* **447**, 465–468 (2004).
8. Kandasamy, P., Gyimesi, G., Kanai, Y. & Hediger, M. A. Amino acid transporters revisited: New views in health and disease. *Trends Biochem. Sci.* (2018). doi:10.1016/j.tibs.2018.05.003
9. Estudante, M. M., Soveral, G., Morais, J. & Benet, L. Insights into Solute Carriers: Physiological Functions and Implications in Disease and Pharmacokinetics. *MedChemComm* (2016). doi:10.1039/C6MD00188B
10. Geier, E. G. *et al.* Profiling Solute Carrier Transporters in the Human Blood–Brain Barrier. *Clin. Pharmacol. Ther.* **94**, 636–639 (2013).

11. Giacomini, K. M. *et al.* International Transporter Consortium Commentary on Clinically Important Transporter Polymorphisms. *Clin. Pharmacol. Ther.* **94**, 23–26 (2013).
12. Clinical Drug Interaction Studies — Study Design, Data Analysis, and Clinical Implications. *Data Anal.* 32
13. Matsson, P. *et al.* Quantifying the impact of transporters on cellular drug permeability. *Trends Pharmacol. Sci.* **36**, 255–262 (2015).
14. Dobson, P. D. & Kell, D. B. Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat. Rev. Drug Discov.* **7**, 205–220 (2008).
15. Mendes, P., Oliver, S. G. & Kell, D. B. Response to 'The Need for Speed', by Matsson et al. *Trends Pharmacol. Sci.* (2016). doi:10.1016/j.tips.2016.02.004
16. Zhao, R., Diop-Bove, N., Visentin, M. & Goldman, I. D. Mechanisms of Membrane Transport of Folates into Cells and Across Epithelia. *Annu. Rev. Nutr.* **31**, 177–201 (2011).
17. Young, J. D., Yao, S. Y. M., Baldwin, J. M., Cass, C. E. & Baldwin, S. A. The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Mol. Aspects Med.* **34**, 529–547 (2013).
18. Bhattacharya, B., Mohd Omar, M. F. & Soong, R. The Warburg effect and drug resistance: The Warburg effect and drug resistance. *Br. J. Pharmacol.* **173**, 970–979 (2016).
19. Icard, P. *et al.* How the Warburg effect supports aggressiveness and drug resistance of cancer cells? *Drug Resist. Updat.* **38**, 1–11 (2018).
20. Winter, G. E. *et al.* The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat. Chem. Biol.* **10**, 768–773 (2014).
21. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* (2015). doi:10.1126/science.aac7557
22. Klein, M.-C. *et al.* AXER is an ATP/ADP exchanger in the membrane of the endoplasmic reticulum. *Nat. Commun.* **9**, 3489 (2018).

23. Robinson, A. J., Kunji, E. R. S. & Gross, A. Mitochondrial carrier homolog 2 (MTCH2): The recruitment and evolution of a mitochondrial carrier protein to a critical player in apoptosis. *Exp. Cell Res.* **318**, 1316–1323 (2012).
24. O'Hagan, S., Wright Muelas, M., Day, P. J., Lundberg, E. & Kell, D. B. GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles. *Cell Syst.* (2018). doi:10.1016/j.cels.2018.01.003
25. Cesar-Razquin, A. *et al.* In silico prioritization of transporter-drug relationships from drug sensitivity screens. (2018). doi:10.1101/381335
26. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
27. Bürckstümmer, T. *et al.* A reversible gene trap collection empowers haploid genetics in human cells. *Nat. Methods* **10**, 965–971 (2013).
28. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* (2017). doi:10.1038/nature22376
29. Nigam, S. K. *et al.* The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective. *Physiol. Rev.* **95**, 83–123 (2015).
30. Licciardello, M. P. *et al.* A combinatorial screen of the CLOUD uncovers a synergy targeting the androgen receptor. *Nat. Chem. Biol.* (2017). doi:10.1038/nchembio.2382
31. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1037
32. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
33. Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* (2016).

34. Sirotnak, F. M., DeGraw, J. I., Moccio, D. M., Samuels, L. L. & Goutas, L. J. New folate analogs of the 10-deaza-aminopterin series Basis for structural design and biochemical and pharmacologic properties. *Cancer Chemother. Pharmacol.* **12**, (1984).
35. Waalkes, T. P. & Makulu, D. R. Pharmacologic aspects of pentamidine. *Natl. Cancer Inst. Monogr.* **43**, 171–177 (1976).
36. Zhang, J. *et al.* The role of nucleoside transporters in cancer chemotherapy with nucleoside drugs. *Cancer Metastasis Rev.* **26**, 85–110 (2007).
37. Hillgren, K. M. *et al.* Emerging Transporters of Clinical Importance: An Update From the International Transporter Consortium. *Clin. Pharmacol. Ther.* **94**, 52–63 (2013).
38. Wittwer, M. B. *et al.* Discovery of Potent, Selective Multidrug and Toxin Extrusion Transporter 1 (MATE1, SLC47A1) Inhibitors Through Prescription Drug Profiling and Computational Modeling. *J. Med. Chem.* **56**, 781–795 (2013).
39. Grottker, J., Rosenberger, A., Burckhardt, G. & Hagos, Y. Interaction of human multidrug and toxin extrusion 1 (MATE1) transporter with antineoplastic agents. *Drug Metabol. Drug Interact.* **26**, (2011).
40. Yang, N.-D. *et al.* Artesunate Induces Cell Death in Human Cancer Cells via Enhancing Lysosomal Function and Lysosomal Degradation of Ferritin. *J. Biol. Chem.* **289**, 33425–33441 (2014).
41. Våtsveen, T. K. *et al.* Artesunate shows potent anti-tumor activity in B-cell lymphoma. *J. Hematol. Oncol.J Hematol Oncol* **11**, (2018).
42. Cui, L. & Su, X. Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Rev. Anti Infect. Ther.* **7**, 999–1013 (2009).
43. Montalbetti, N., Simonin, A., Kovacs, G. & Hediger, M. A. Mammalian iron transporters: Families SLC11 and SLC40. *Mol. Aspects Med.* **34**, 270–287 (2013).
44. Pérez-Escuredo, J. *et al.* Monocarboxylate transporters in the brain and in cancer. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1863**, 2481–2497 (2016).

45. Song, Z. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Mol. Aspects Med.* **34**, 590–600 (2013).
46. Schiöth, H. B., Roshanbin, S., Hägglund, M. G. A. & Fredriksson, R. Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects. *Mol. Aspects Med.* **34**, 571–585 (2013).
47. Bigenzahn, J. W. *et al.* LZTR1 is a regulator of RAS ubiquitination and signaling. *Science* eaap8210 (2018). doi:10.1126/science.aap8210
48. O'Shea, E. K. *et al.* Identification and characterization of a transporter complex responsible for the cytosolic entry of nitrogen-containing bisphosphonates. (2018). doi:10.1101/289835
49. Wenlock, M. C., Austin, R. P., Barton, P., Davis, A. M. & Leeson, P. D. A Comparison of Physiochemical Property Profiles of Development and Marketed Oral Drugs. *J. Med. Chem.* **46**, 1250–1256 (2003).
50. Lai, H. & Singh, N. P. Selective cancer cell cytotoxicity from exposure to dihydroartemisinin and holotransferrin. *Cancer Lett.* **91**, 41–46 (1995).
51. Duan, G. *et al.* Increased Glutamine Consumption in Cisplatin-Resistant Cells Has a Negative Impact on Cell Growth. *Sci. Rep.* **8**, (2018).
52. Guidi, N. & Longo, V. D. Periodic fasting starves cisplatin-resistant cancers to death. *EMBO J.* **37**, e99815 (2018).
53. Farge, T. *et al.* Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. *Cancer Discov.* **7**, 716–735 (2017).
54. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
55. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).

56. Stransky, N. *et al.* Pharmacogenomic agreement between two cancer cell line data sets. *Nature* (2015). doi:10.1038/nature15736
57. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
58. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
59. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
60. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
61. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17224—17229 (2014).
62. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nat. Methods* **11**, 122–123 (2014).
63. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
64. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* (2014). doi:10.1038/nature14136
65. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
66. Berthold, M. R. *et al. Technical Report Knime: The Konstanz Information Miner*.
67. *RDKit: Open-source cheminformatics*.
68. Swain, M. *MolVS: Molecule Validation and Standardization*. (20017–).
69. Python Core Team. *Python : A dynamic, open source programming language*. (Python Software Foundation, 2018).

70. RStudio Team. *RStudio: Integrated Development Environment for R*. (RStudio, Inc., 2015).
71. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2018).
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
73. Wei, T. & Simko, V. *R package 'corrplot': Visualization of a Correlation Matrix*. (2017).
74. Kassambara, A. & Mundt, F. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. (2017).
75. Lê, S., Josse, J. & Husson, F. FactoMineR: A Package for Multivariate Analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
76. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

Figure legends

Figure 1. a: Schematic view of the composition of the SLC-focused CRISPR/Cas9 library and experimental outline of the genetic screen. **b:** Circular plot showing SLCs expressed in Hap1 cells according to RNAseq data from Brockmann *et al.* SLC families are indicated in the inner circle while expression level (log_2 counts per 10^7 reads) is shown as blue bars. SLCs with an expression level above 9 are labeled. **c:** Schematic view of the compound sets and steps applied to the selection of a final set of drugs for screening. **D:** Principal component analysis of compounds in the DrugBank set of reference as well as in the sets tested in this study based on 22 annotated 2D chemical descriptors.

Figure 2. a: sgRNA-level enrichment for samples treated with 10X IC_{50} methotrexate, as determined by DESeq2. All six sgRNAs targeting the SLC19A1 gene show significant enrichment. **b:** Gene-level enrichment for samples treated with 10X IC_{50} methotrexate, as determined by GSEA. Average log2 fold change for the significant sgRNAs for each gene is shown in the x-axis. **c:** Overview of significantly enriched SLCs (FDR<1%) identified upon treatment with different compounds. Significant enrichments for all different doses of the same compound are merged together in order to ease interpretation (union), always selecting the most significant value for repeats. SLC genes are ordered by name, and treatments are ordered by hierarchical clustering based on the gene-level results.

Figure 3: a: Schematic view of the Multicolor Competition Assay (MCA). **b:** Validation of selected SLC/drug associations by MCA. Results are shown by gene tested. Ratios of GFP+/mCherry+ populations are shown for the indicated SLC/drug combinations at the given timepoints.

Supplemental Figure 1: a: Violin plots of sgRNA count distributions in the SLC library plasmid samples and in the 9 days post-infection (p.i.) samples . **b:** Volcano plot (p-value vs. log2 fold change) for the differential representation of sgRNAs in samples collected 9 days p.i. vs. the plasmid library. P-values correspond to a two-tailed Wald test (DESeq2). sgRNA

corresponding to the set of 20 essential control genes are shown in green. **c:** Same as b., but in this case sgRNAs corresponding to the set of 120 non-target negative control sequences are shown in red. **d:** Gene-level enrichment in 9 days p.i. vs. plasmid library. **e:** YM155 benchmarking screen. Read counts for the samples at day0, DMSO-treated and YM155-treated samples are shown. **f:** Venn diagram showing the compound subsets used for the chemoinformatic analysis **g:** Correlogram plot showing the 2D descriptors contribution to the PCA analysis.

Suppl Figure 2: a: Overview of significantly enriched SLCs (FDR<1%) identified for all drug treatments. **b:** Expression levels in HAP1 cells for SLCs significantly enriched in our screen.

Figure 1

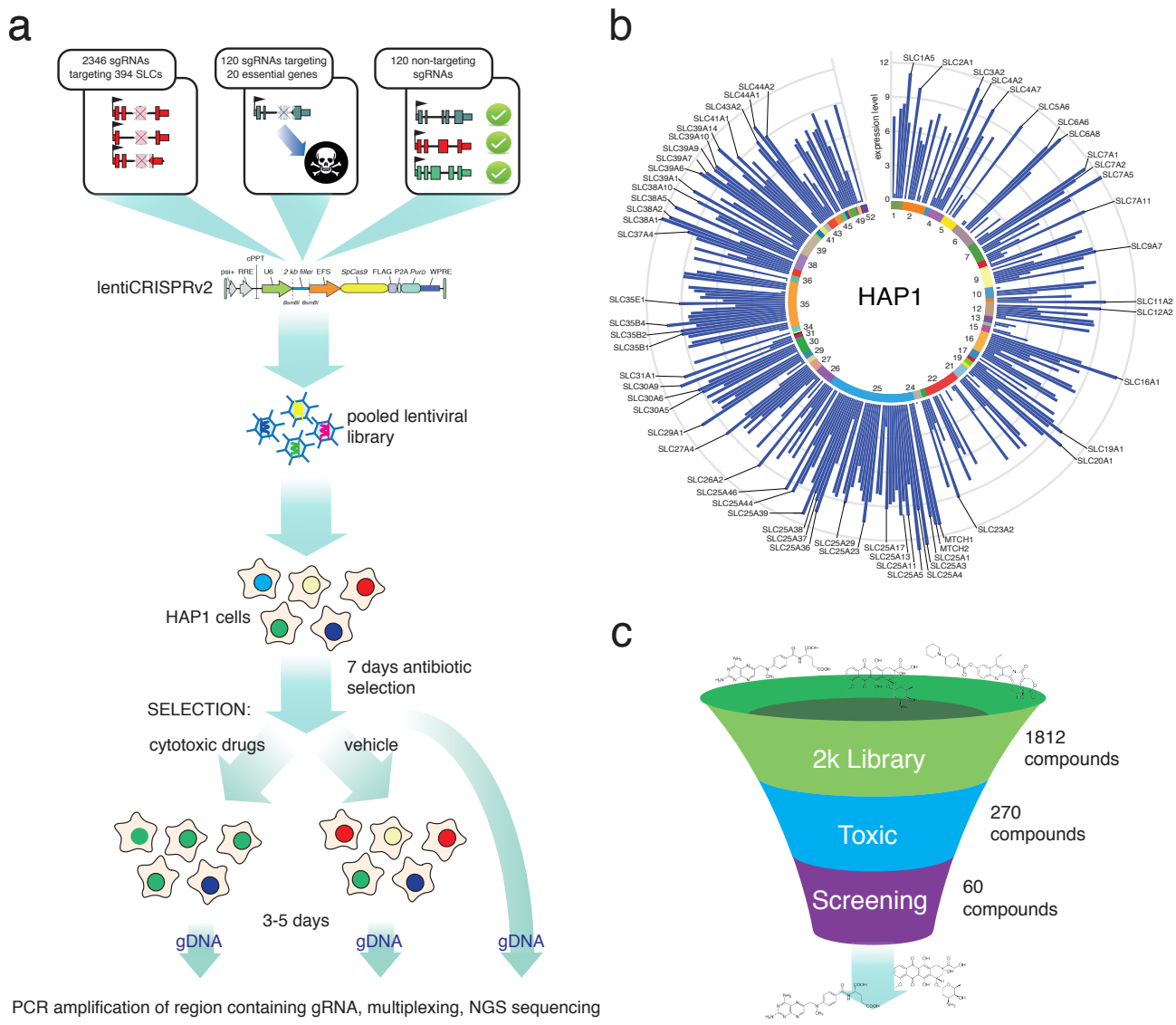

a
2346 sgRNAs targeting 394 SLCs
120 sgRNAs targeting 20 essential genes
120 non-targeting sgRNAs
lentiCRISPRv2
pooled lentiviral library
HAP1 cells
7 days antibiotic selection
SELECTION:
cytotoxic drugs
vehicle
3-5 days
gDNA
gDNA
gDNA
PCR amplification of region containing gRNA, multiplexing, NGS sequencing
b
HAP1
expression level
c
2k Library
1812 compounds
Toxic
270 compounds
Screening
60 compounds
d
Subsets
Drugbank
2k Library
Toxic
Screening
PC1 (44.9%)
PC2 (17.2%)

Figure 2

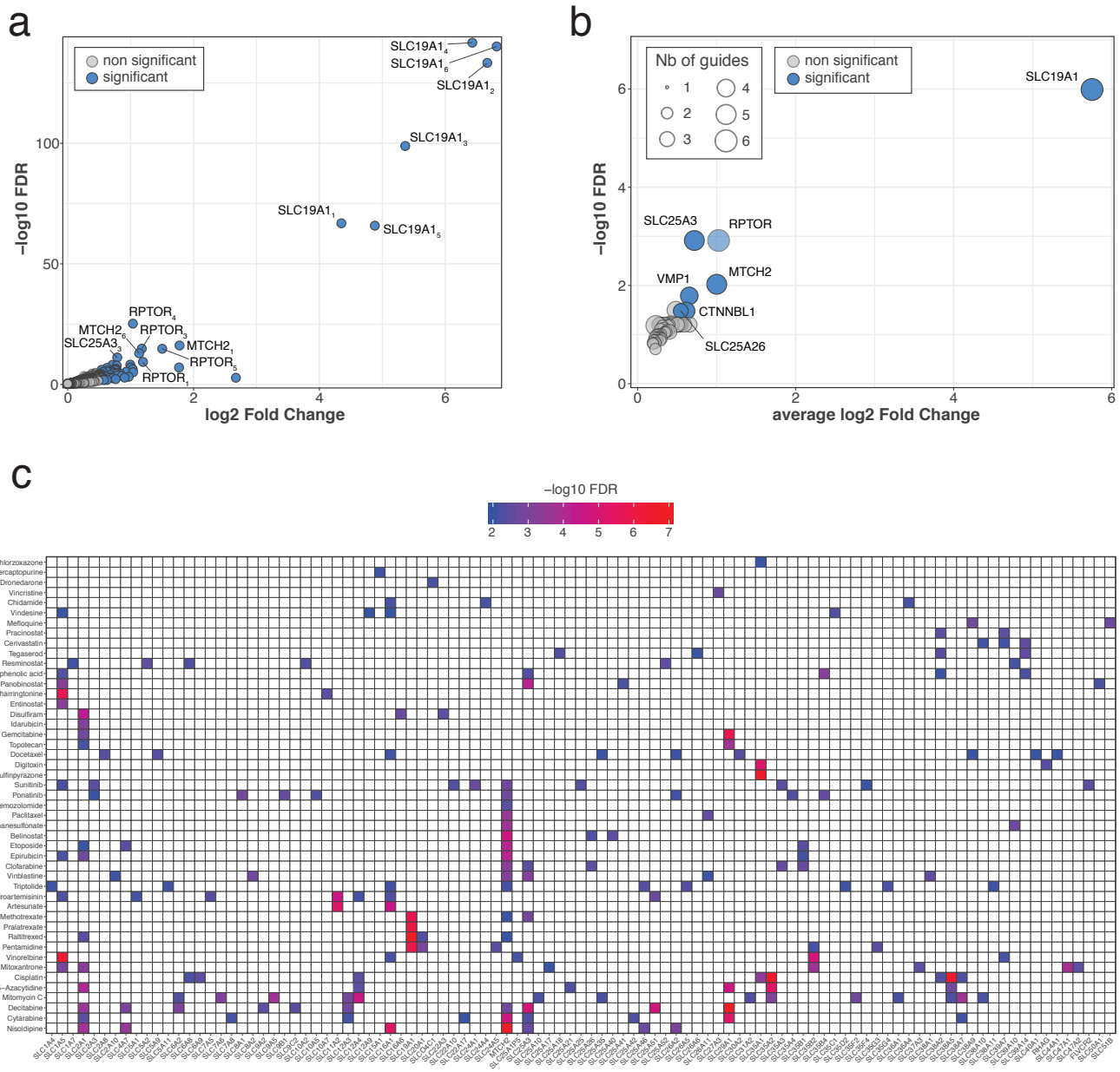
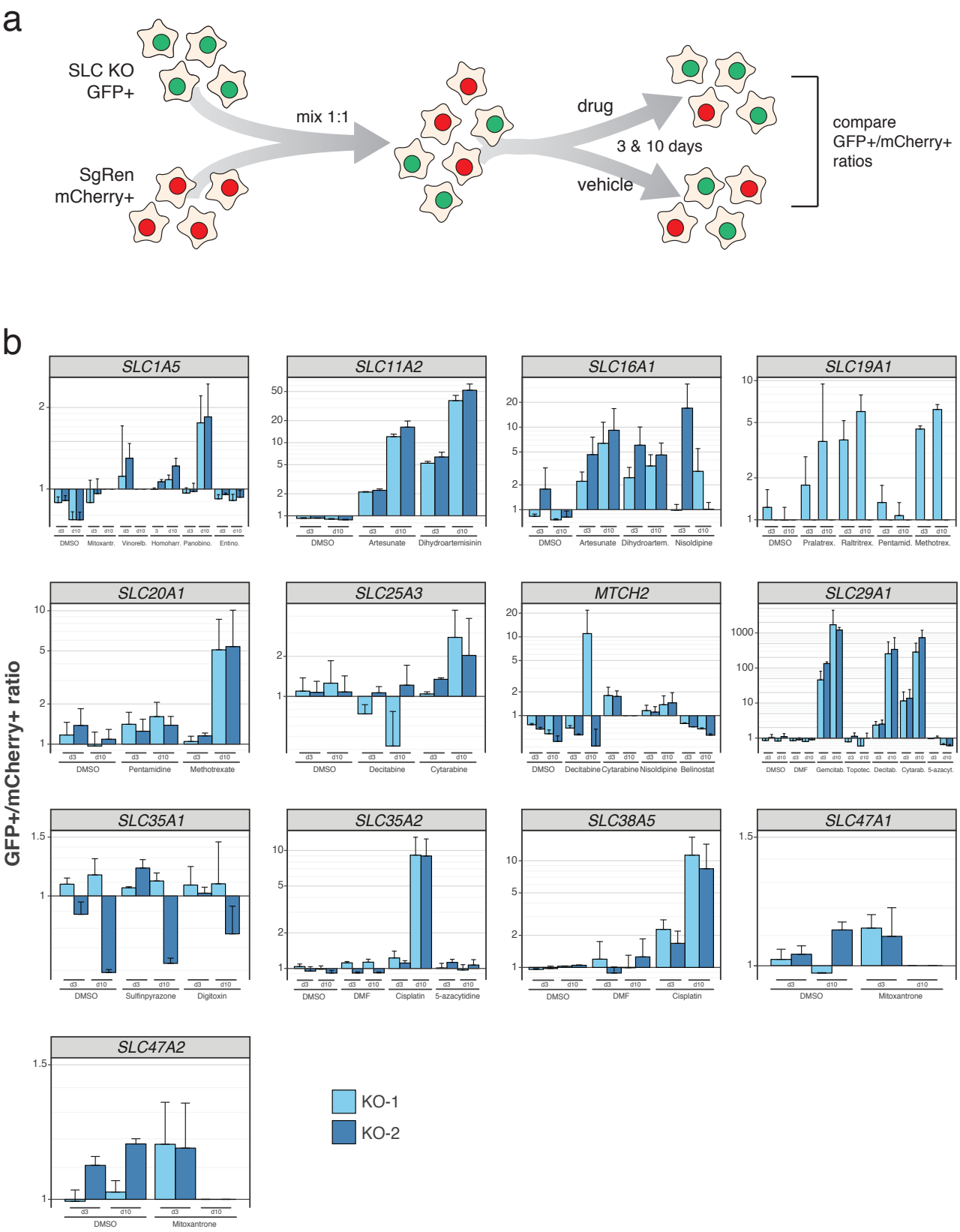
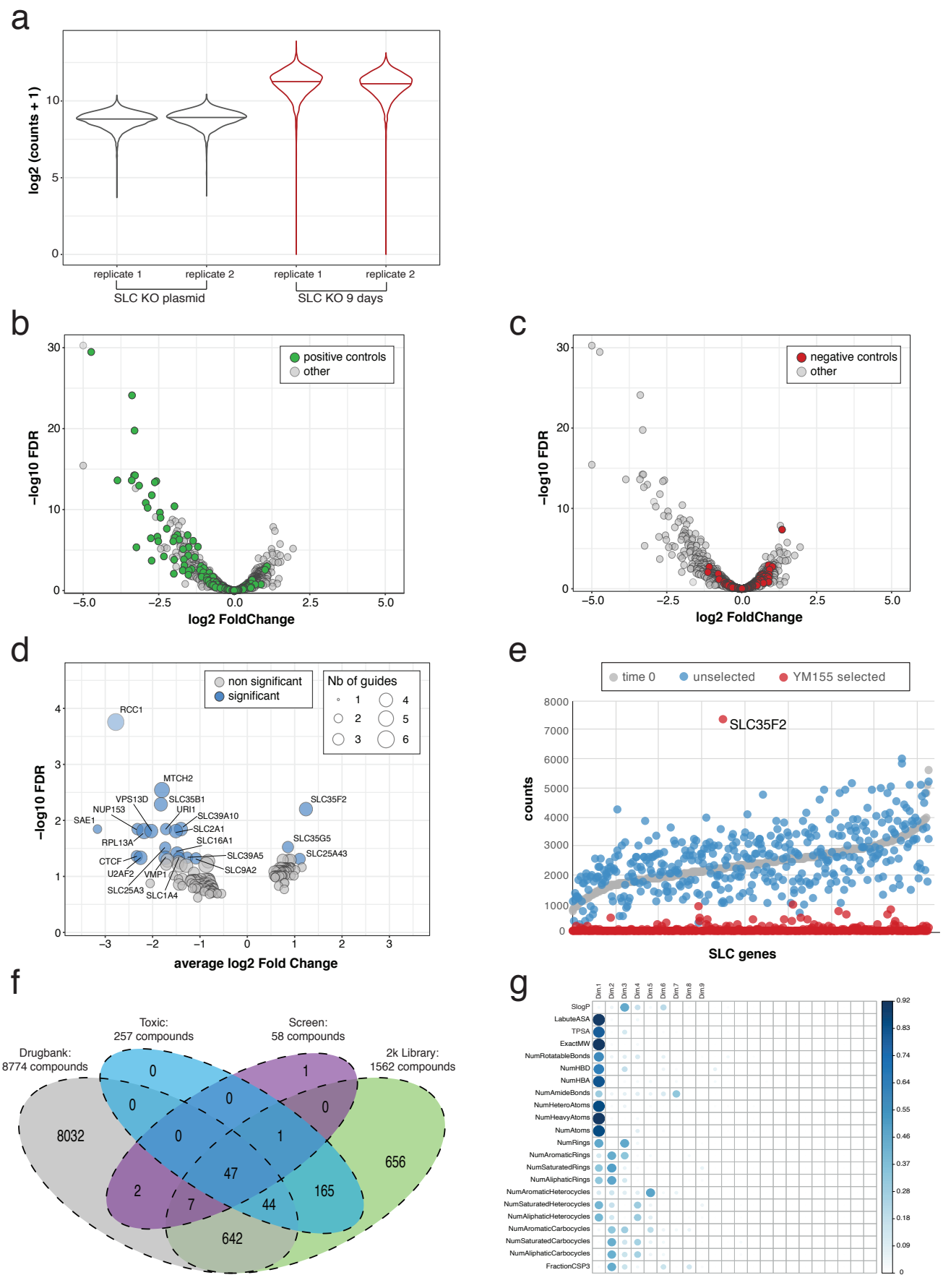

a
non significant
significant
SLC19A1 4
SLC19A1 6
SLC19A1 2
SLC19A1 3
SLC19A1 1
SLC19A1 5
RPTOR 4
MTCH2 6
RPTOR 3
SLC25A3 3
MTCH2 1
RPTOR 5
RPTOR 1
−log10 FDR
log2 Fold Change
b
Nb of guides
1
2
3
4
5
6
non significant
significant
SLC19A1
SLC25A3
RPTOR
MTCH2
VMP1
CTNNBL1
SLC25A26
−log10 FDR
average log2 Fold Change
c
−log10 FDR
2 3 4 5 6 7
Chlorzoxazone
6-mercaptopurine
Dronedarone
Vincristine
Chidamide
Vindesine
Mefloquine
Pracinostat
Cerivastatin
Tegaserod
Resminostat
Mycophenolic acid
Panobinostat
Homoharringtonine
Entinostat
Disulfiram
Idarubicin
Gemcitabine
Topotecan
Docetaxel
Digitoxin
Sulfinpyrazone
Sunitinib
Ponatinib
Temozolomide
Paclitaxel
Methyl methanesulfonate
Belinostat
Etoposide
Epirubicin
Clofarabine
Vinblastine
Triptolide
Dihydroartemisinin
Artesunate
Methotrexate
Pralatrexate
Raltitrexed
Pentamidine
Vinorelbine
Mitoxantrone
Cisplatin
5−Azacytidine
Mitomycin C
Decitabine
Cytarabine
Nisoldipine

Figure 3

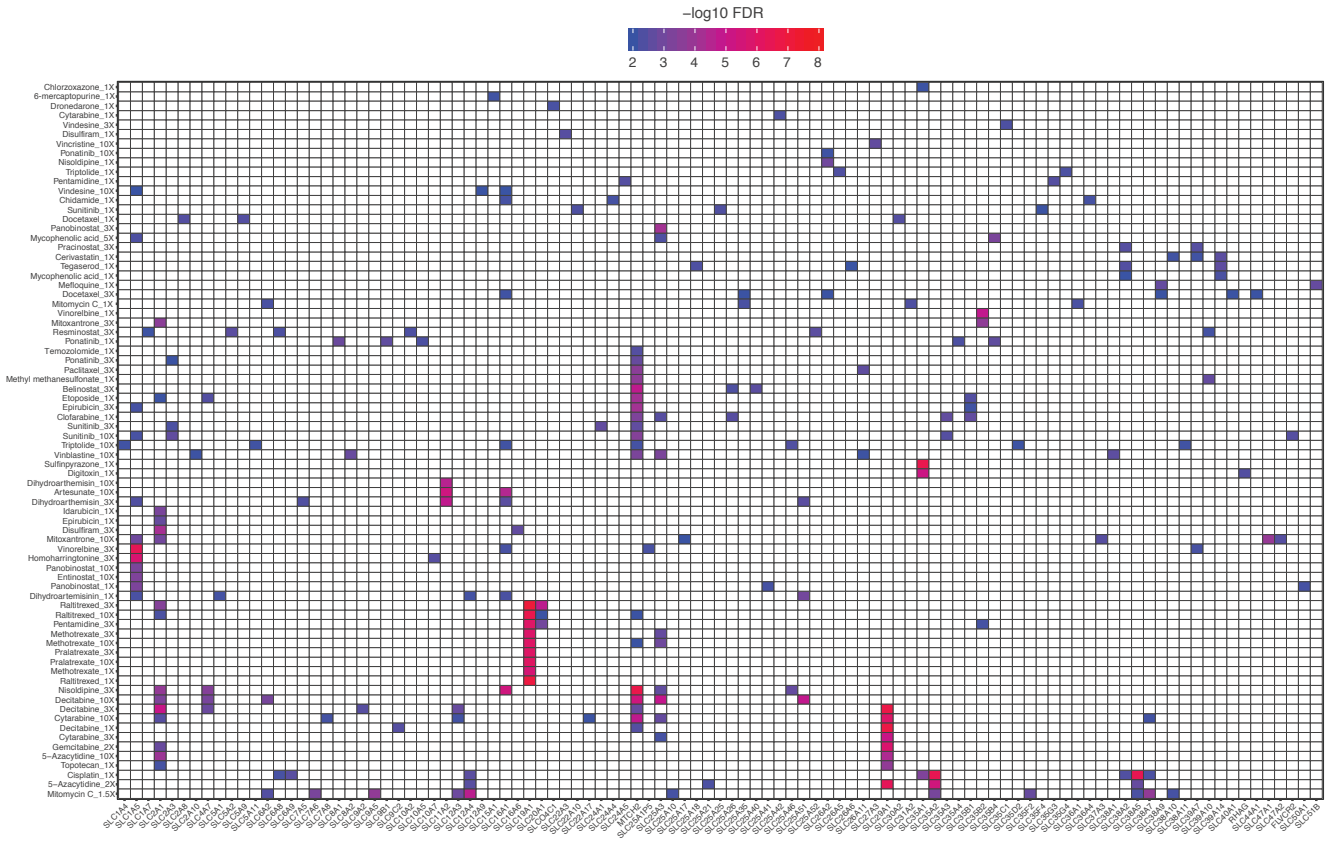

a
SLC KO
GFP+
SgRen
mCherry+
mix 1:1
drug
vehicle
3 & 10 days
compare
GFP+/mCherry+
ratios
b
GFP+/mCherry+ ratio
SLC1A5
DMSO
Mitoxantr.
Vinorelb.
Homoharr.
Panobino.
Entino.
SLC11A2
DMSO
Artesunate
Dihydroartemisinin
SLC16A1
DMSO
Artesunate
Dihydroartem.
Nisoldipine
SLC19A1
DMSO
Pralatrex.
Raltritrex.
Pentamid.
Methotrex.
SLC20A1
DMSO
Pentamidine
Methotrexate
SLC25A3
DMSO
Decitabine
Cytarabine
MTCH2
DMSO
Decitabine
Cytarabine
Nisoldipine
Belinostat
SLC29A1
DMSO
DMF
Gemcitab.
Topotec.
Decitab.
Cytarab.
5-azacyt.
SLC35A1
DMSO
Sulfinpyrazone
Digitoxin
SLC35A2
DMSO
DMF
Cisplatin
5-azacytidine
SLC38A5
DMSO
DMF
Cisplatin
SLC47A1
DMSO
Mitoxantrone
SLC47A2
DMSO
Mitoxantrone
KO-1
KO-2

Supplemental Figure 1

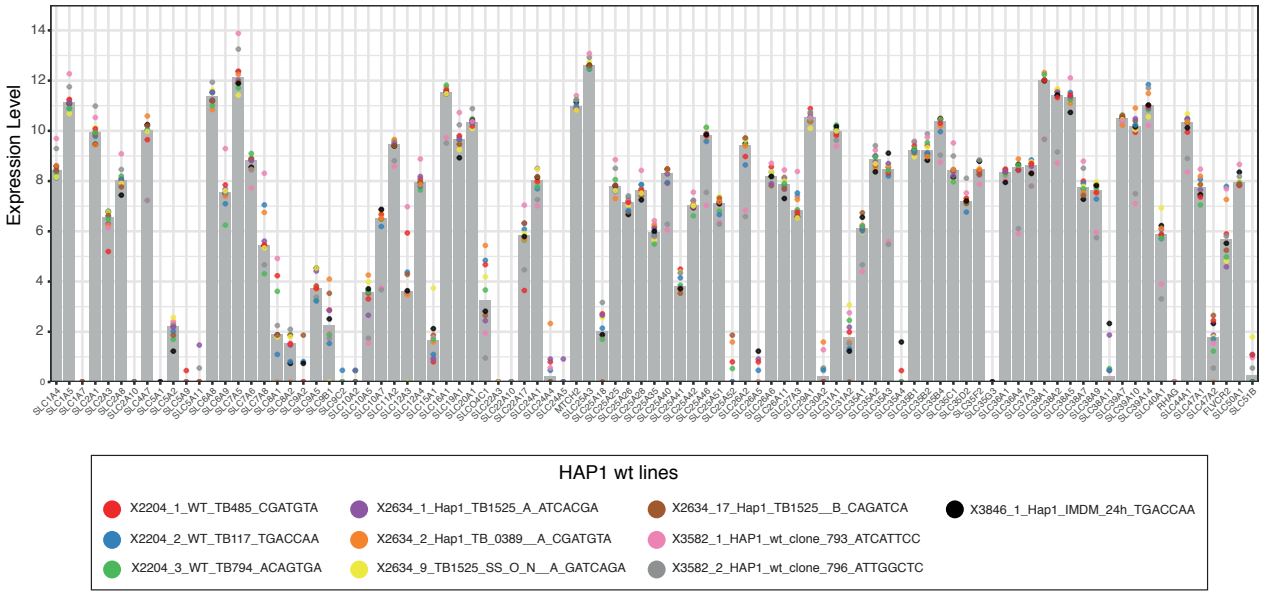


Supplemental Figure 2

a



b



Supplemental Table 1. Targeted SLCs and positive controls (~6 sgRNAs per gene)

SLC genes and pseudogenes									Positive controls
SLC1A1	SLC6A1	*SLC9A7P1*	SLC16A10	SLC22A12	SLC25A26	SLC27A6	SLC35F2	SLC41A1	CEP85
SLC1A2	SLC6A2	SLC9A8	SLC16A11	SLC22A13	SLC25A27	SLC28A1	SLC35F3	SLC41A2	CTCF
SLC1A3	SLC6A3	SLC9A9	SLC16A12	SLC22A14	SLC25A28	SLC28A2	SLC35F4	SLC41A3	CTNNBL1
SLC1A4	SLC6A4	SLC9B1	SLC16A13	SLC22A15	SLC25A29	SLC28A3	SLC35F5	RHAG	DYRK1A
SLC1A5	SLC6A5	SLC9B2	SLC16A14	SLC22A16	SLC25A30	SLC29A1	SLC35F6	RHBG	KANSL1
SLC1A6	SLC6A6	SLC9C1	SLC17A1	SLC22A17	SLC25A31	SLC29A2	SLC35G1	RHCG	MED26
SLC1A7	SLC6A7	SLC9C2	SLC17A2	SLC22A18	SLC25A32	SLC29A3	SLC35G2	SLC43A1	NBAS
SLC2A1	SLC6A8	SLC10A1	SLC17A3	SLC22A20	SLC25A33	SLC29A4	SLC35G3	SLC43A2	NFYC
SLC2A2	SLC6A9	SLC10A2	SLC17A4	SLC22A23	SLC25A34	SLC30A1	SLC35G4	SLC43A3	NIPBL
SLC2A3	SLC6A11	SLC10A3	SLC17A5	SLC22A24	SLC25A35	SLC30A2	SLC35G5	SLC44A1	NUP153
SLC2A3P1	SLC6A12	SLC10A4	SLC17A6	SLC22A25	SLC25A36	SLC30A3	SLC35G6	SLC44A2	PAFAH1B1
SLC2A4	SLC6A13	SLC10A5	SLC17A7	SLC22A31	SLC25A37	SLC30A4	SLC36A1	SLC44A3	RCC1
SLC2A5	SLC6A14	SLC10A6	SLC17A8	SLC23A1	SLC25A38	SLC30A5	SLC36A2	SLC44A4	RPL13A
SLC2A6	SLC6A15	SLC10A7	SLC17A9	SLC23A2	SLC25A39	SLC30A6	SLC36A3	SLC44A5	RPTOR
SLC2A7	SLC6A16	SLC11A1	SLC18A1	SLC23A3	SLC25A40	SLC30A7	SLC36A4	SLC45A1	SAE1
SLC2A8	SLC6A17	SLC11A2	SLC18A2	SLC24A1	SLC25A41	SLC30A8	SLC37A1	SLC45A2	SRBD1
SLC2A9	SLC6A18	SLC12A1	SLC18A3	SLC24A2	SLC25A42	SLC30A9	SLC37A2	SLC45A3	U2AF2
SLC2A10	SLC6A19	SLC12A2	SLC18B1	SLC24A3	SLC25A43	SLC30A10	SLC37A3	SLC45A4	URI1
SLC2A11	SLC6A20	SLC12A3	SLC19A1	SLC24A4	SLC25A44	SLC31A1	SLC37A4	SLC46A1	VMP1
SLC2A12	SLC7A1	SLC12A4	SLC19A2	SLC24A5	SLC25A45	SLC31A2	SLC38A1	SLC46A2	VPS13D
SLC2A13	SLC7A2	SLC12A5	SLC19A3	MTCH1	SLC25A46	SLC32A1	SLC38A2	SLC46A3	
SLC2A14	SLC7A3	SLC12A6	SLC20A1	MTCH2	SLC25A47	SLC33A1	SLC38A3	SLC47A1	
SLC3A1	SLC7A4	SLC12A7	SLC20A2	SLC25A1	SLC25A48	SLC34A1	SLC38A4	SLC47A2	
SLC3A2	SLC7A5	SLC12A8	SLCO1A2	*SLC25A1P5*	SLC25A51	SLC34A2	SLC38A5	SLC48A1	
SLC4A2	*SLC7A5P1*	SLC12A9	SLCO1B1	SLC25A2	SLC25A52	SLC34A3	SLC38A6	DIRC2	
SLC4A3	*SLC7A5P2*	SLC13A1	SLCO1B3	SLC25A3	SLC25A53	SLC35A1	SLC38A7	FLVCR1	
SLC4A4	SLC7A6	SLC13A2	SLCO1C1	SLC25A4	UCP1	SLC35A2	SLC38A8	FLVCR2	
SLC4A5	SLC7A7	SLC13A3	SLCO2A1	SLC25A5	UCP2	SLC35A3	SLC38A9	MFSD7	
SLC4A7	SLC7A8	SLC13A4	SLCO2B1	SLC25A6	UCP3	SLC35A4	SLC38A10	SLC50A1	
SLC4A8	SLC7A9	SLC13A5	SLCO3A1	SLC25A10	SLC26A1	SLC35A5	SLC38A11	SLC51A	
SLC4A9	SLC7A10	SLC14A1	SLCO4A1	SLC25A11	SLC26A2	SLC35B1	SLC39A1	SLC51B	
SLC4A10	SLC7A11	SLC14A2	SLCO4C1	SLC25A12	SLC26A3	SLC35B2	SLC39A2	SLC52A1	
SLC4A11	SLC7A13	SLC15A1	SLCO5A1	SLC25A13	SLC26A4	SLC35B3	SLC39A3	SLC52A2	
SLC5A1	SLC7A14	SLC15A2	SLCO6A1	SLC25A14	SLC26A5	SLC35B4	SLC39A4	SLC52A3	
SLC5A2	SLC8A1	SLC15A3	SLC22A1	SLC25A15	SLC26A6	SLC35C1	SLC39A5		
SLC5A3	SLC8A2	SLC15A4	SLC22A2	SLC25A16	SLC26A7	SLC35C2	SLC39A6		
SLC5A4	SLC8A3	SLC16A1	SLC22A3	SLC25A17	SLC26A8	SLC35D1	SLC39A7		
SLC5A5	SLC8B1	SLC16A2	SLC22A4	SLC25A18	SLC26A9	SLC35D2	SLC39A8		
SLC5A6	SLC9A1	SLC16A3	SLC22A5	SLC25A19	SLC26A10	SLC35D3	SLC39A9		
SLC5A7	SLC9A2	SLC16A4	SLC22A6	SLC25A20	SLC26A11	SLC35E1	SLC39A10		
SLC5A8	SLC9A3	SLC16A5	SLC22A7	SLC25A21	SLC27A1	*SLC35E1P1*	SLC39A11		
SLC5A9	SLC9A4	SLC16A6	SLC22A8	SLC25A22	SLC27A2	SLC35E2B	SLC39A12		
SLC5A10	SLC9A5	SLC16A7	SLC22A9	SLC25A23	SLC27A3	SLC35E3	SLC39A13		
SLC5A11	SLC9A6	SLC16A8	SLC22A10	SLC25A24	SLC27A4	SLC35E4	SLC39A14		
SLC5A12	SLC9A7	SLC16A9	SLC22A11	SLC25A25	SLC27A5	SLC35F1	SLC40A1		

Supplemental Table 2. Screened compounds.

Class	Subclass	Name	Status*
antineoplastic	purine analogs	6-mercaptopurine	A
	nucleoside analogs	5-azacitidine	A
		clofarabine	A
		cytarabine	A
		decitabine	A
		gemcitabine	A
	antifolates	methotrexate	A
		pralatrexate	A
		raltitrexed	A
	HDAC inhibitors	belinostat	A
		chidamide (tucidinostat)	I
		entinostat	I
		panobinostat	A
		pracinostat	I
		resminostat	I
		romidepsin	A
		vorinostat	A
	microtubule inhibitors (destabilizing)	vinblastine	A
		vincristine	A
		vindesine	A
		vinorelbine	A
	microtubule inhibitors (stabilizing)	docetaxel	A
		paclitaxel	A
	proteasome inhibitors	bortezomib	A
		carfilzomib	A
	RTK inhibitors	crizotinib	A
		ponatinib	A
		sunitinib	A
	topoisomerase I inhibitors	topotecan	A
		irinotecan	A

*based on DrugBank: A (approved), A, W (approved, withdrawn), I (investigational), E (experimental)

Supplemental Table 2 (cont). Screened compounds.

Class	Subclass	Name	Status*
antineoplastic	topoisomerase II inhibitors	doxorubicin	A
		epirubicin	A
		etoposide	A
		idarubicin	A
		mitoxantrone	A
	protein translation inhibitors	homoharringtonine (omacetaxine mepesuccinate)	A
	transcription inhibitors	dactinomycin	A
	alkylating	cisplatin	A
		methyl methanesulfonate	A
		mitomycin C	A
		temozolomide	A
	other	triptolide	I
antiparasitic	antimalarial	artesunate	A
		dihydroartemisinin (artenimol)	I
		mefloquine	A
	antihelmintic	albendazole	A
	antiprotozoal	pentamidine	A
antiarrhythmic	type III: K-channel blocker	amiodarone	A
		dronedarone	A
	type V	digitoxin	A
antihypertensive	Ca-channel blocker	nisoldipine	A
anti-inflammatory	NSAID	oxyphenbutazone	A, W
immunosuppressant		mycophenolic acid	A
hypolipidemic	HMG-CoA reductase inhibitor	cerivastatin	A, W
antipasmodic		chlorzoxazone	A
		metaxalone	A
prokinetic	serotonin agonist	tegaserod	A, W
mineralocorticoid		desoxycorticosterone pivalate	E
uricosuric		sulfinpyrazone	A
alcohol deterrent		disulfiram	A

*based on DrugBank: A (approved), A, W (approved, withdrawn), I (investigational), E (experimental)

Supplemental Table 3. Descriptors used in the chemical space analysis

Descriptor name	Description
SlogP	Smarts LogP, Octanol Water Partition Coefficient
LabuteASA	Labute's Approximate Surface Area, approximated surface area of a molecule (J. Mol. Graph. Mod. 18, 464-77 (2000))
TPSA	Total Polar surface area
ExactMW	Molecular weight
NumRotatableBonds	Number of rotatable bonds
NumHBD	Number of hydrogen bond donors
NumHBA	Number of hydrogen bond acceptors
NumAmideBonds	Number of amide bonds
NumHeteroAtoms	Number of hetero atoms
NumHeavyAtoms	Number of heavy atoms
NumAtoms	Number of atoms
NumRings	Number of rings
NumAromaticRings	Number of aromatic rings
NumSaturatedRings	Number of saturated rings
NumAliphaticRings	Number of aliphatic rings
NumAromaticHeterocycles	Number of aromatic heterocycles
NumSaturatedHeterocycles	Number of saturated heterocycles
NumAliphaticHeterocycles	Number of aliphatic heterocycles
NumAromaticCarbocycles	Number of aromatic carbocycles
NumSaturatedCarbocycles	Number of saturated carbocycles
NumAliphaticCarbocycles	Number of aliphatic carbocycles
FractionCSP3	Fraction of sp3 hybridized Carbons

Supplemental Table 4. Selected SLC-drug interactions for validation

Gene	Drug
SLC1A5	Mitoxantrone
	Vinorelbine
	Homoharringtonine
	Panobinostat
	Entinostat
SLC11A2	Artesunate
	Dihydroarthemisinin
SLC16A1	Artesunate
	Dihydroarthemisinin
	Nisoldipine
SLC19A1	Pralatraxate
	Raltitrexed
	Pentamidine
	Methotrexate
SLC20A1	Pentamidine
	Methotrexate
SLC25A3	Decitabine
	Cytarabine
MTCH2	Decitabine
	Cytarabine
	Nisoldipine
	Belinostat
SLC29A1	Gemcitabine
	Topotecan
	Decitabine
	Cytarabine
	5-azacytidine
SLC35A1	Sulfinpyrazone
	Digitoxin
SLC35A2	Cisplatin
	5-azacytidine
SLC38A5	Cisplatin
SLC47A1	Mitoxantrone
SLC47A2	Mitoxantrone

3 Discussion

3.1 General discussion

Transmembrane transporters are known mediators of drug uptake and disposition, but the extent of their contribution is only partially understood. While it has been traditionally assumed that drug-like compounds can passively diffuse across lipoidal membranes following the well-known Lipiniski's "rule of 5" (Ro5), the hypothesis that every drug needs a specific carrier to enter cells has been gaining support in the last years, although it still remains a matter of controversy (Di et al, 2012; Dobson & Kell, 2008; Kell, 2015; 2016; Matsson et al, 2015; Smith et al, 2014; Sugano et al, 2010). If that were the case, new dependencies between drugs and transporters are to be discovered that could shed light on the pharmacokinetics and pharmacodynamics of many therapeutic agents.

Of the two main protein superfamilies involved in drug transport, the ATP-binding cassette (ABC) transporters are relatively well characterized and many of its members are known promiscuous carriers that mediate the export of a great deal of compounds and are involved in multidrug resistance in cancer (MDR). In contrast, the larger and heterogenous group of Solute carriers (SLC), with more than 400 members, does not seem to garner the attention it deserves given its relevance for cell physiology and pharmacology.

Indeed, we have shown that SLCs have the largest publication asymmetry of all human gene groups as well as a lower number of publications than the average. This effect cannot be explained solely by their transmembrane nature, which makes them generally more difficult to isolate and characterize than soluble proteins, since other membrane protein families such as ABCs, ion channels or GPCRs do not show such uneven distributions. Furthermore, the identity of the SLCs that receive most of the research attention seems to have changed only very slightly in the last decade, pointing to a lack of exploratory efforts within the family. Together with technical barriers for the establishment of cell systems and screening assays, the late adoption of a unifying nomenclature for SLCs, which has obscured the consideration

of common principles, and the overlapping specificities and compensatory effects of many of these transporters, which can confound experimental results, might have strongly contributed to this knowledge deficit (Hediger et al, 2013; Hediger et al, 2004).

More specifically, SLCs also seem to be understudied from the standpoint of drug discovery. Although there is ample evidence of their druggability, only ~7% of SLCs (~26 members) are targets of known drugs or drugs in development. This is clearly far from the hundreds of GPCR targets that have already been established, which account for at least 40% of this family (excluding olfactory receptors) (Lin et al, 2015; Rask-Andersen et al, 2013; Rask-Andersen et al, 2014). When considering SLCs associated with human diseases, only one fourth was found to be targeted by at least one active compound, therefore leaving an enormous underexplored space of great therapeutic potential. Similarly, a compilation of currently known drug-transporter interactions revealed that most of the cases involved only a few SLC families of broad specificity (i.e. SLCO, SLC22, SLC15, SLC47), although transport relationships with carriers of narrower specificities are also being increasingly described (Fets et al, 2018; Winter et al, 2014; Yu et al, 2018; Zhao et al, 2011). Therefore, given their undercharacterized or even orphan nature as well as the known participation of some of their members in drug transport, SLCs seem to be promising candidates for unraveling the transport mechanism and metabolic dependencies of many drugs and drug-like compounds.

Within the framework of this thesis, we have aimed at identifying these SLC-drug relationships by using both experimental and computational methods in two orthogonal approaches, namely drug sensitivity predictive modeling and CRISPR-Cas9 genetic screening.

Predictive models of drug sensitivity built on pharmacogenomics datasets of panels of molecularly annotated cancer cell lines have been extensively used in order to find genetic markers that affect drug response as well as to uncover the mechanism of action of drug-like compounds (Barretina et al, 2012; Garnett et al, 2012; Iorio et al, 2016). In the case of transporters, previous studies have already sought to associate drug response with the expression level of members of the ABC and SLCO/SLC22 families (Okabe et al, 2008; Szakács et al, 2004). However, these studies were based on a relatively limited set of cell lines (the NCI-60 panel) and only looked for correlations with individual genes, therefore possibly overlooking

the redundancy and cooperative effects that often exist among transporters. Moreover, they only considered the two best known SLC families involved in drug transport, ignoring all the ~400 remaining members of this group. In contrast, we now exploited a more recent and comprehensive pharmacogenomics dataset while considering the combined effect of all transporters at once in linear models of drug sensitivity.

Regularized linear regression is a powerful method that has already been successfully used for the prediction of drug sensitivity, although generally in a genome-wide context (Barretina et al, 2012; Garnett et al, 2012; Papillon-Cavanagh et al, 2013). One of the main strengths of these techniques, apart from their robustness, is the interpretability of their results. By using large numbers of predictors, they achieve high predictive powers, while the simple, weighted linear combination of features in their models allows the identification and assessment of individual contributions. In our case, we used SLC and ABC molecular data only, with the aim of prioritizing drugs that show a greater dependency on transporters and identifying specific individual transporter associations to sensitive and resistant phenotypes, thereby removing the effect of other genes. This is, to our knowledge, the first application of regularized linear regression to this specific set of proteins.

Of the different data types employed for prediction, gene expression alone turned out to be the most predictive, even superior to the combination with genomic descriptors. This was in agreement with previous studies, and it is likely due to the increased information content present in continuous expression values compared with the discrete nature of copy number variations (CNVs) or single nucleotide variants (SNVs) (Geeleher et al, 2014; Jang et al, 2014). Using gene expression data for modeling, we obtained good predictive models (CI$\geq$0.6) for more than half of the compounds included in the dataset, and then ranked genes within the models by frequency of appearance and relative weight. This allowed us to identify both known and undescribed individual transporter-drug relationships as well as cooperative effects. For instance, drug response to the anticancer drug sepantronium bromide (YM155) was mainly determined by the expression level of two transporters acting in opposite directions, which corresponded to its already described uptake and extrusion carriers (SLC35F2 and ABCB1, respectively). Also, resistance associations with the promiscuous exporter ABCB1 included known substrates (e.g. tanespimycin, vinblastine, docetaxel) as well as other compounds not

previously reported (e.g. ZG-10, PHA-793887, WZ3105). More interestingly, the group of MEK inhibitors, for which no specific transporters have been yet reported, displayed a similar pattern of sensitivity associations to several genes. Such relationships point to cooperative interactions among transporters that in some cases are more likely to consist in indirect effects that affect cell metabolism (e.g. the melanoma-associated SLC45A2 or the sodium-dependent phosphate transporter SLC20A1) while in other could correspond to actual transport mechanisms (e.g. the orphan member of the SLC22 family of drug transporters SLC22A15).

Experimental validation will be required to confirm such associations as well as to distinguish direct from indirect effects. To this end, specific transport assays and drug sensitivity testing in cells with altered transporter expression, such as CRISPR-Cas9 knock-outs or RNAi knock-downs, could be used. As an illustration, a recent study published after our Elastic Net analysis confirmed SLC16A7 (MCT2) as the main uptake transporter of the HIF prolyl-hydroxylase inhibitor dimethyloxalylglycin (DMOG) (Fets et al, 2018). DMOG was indeed one of the compounds for which we obtained better predictive models (CI=0.67) and SLC16A7 appeared as its main sensitivity predictor. Similarly, another member of the same family, SLC16A1 (MCT1), ranked second in our analysis and was also proven to mediate DMOG transport, although with a much lower affinity than SLC16A7. Nevertheless, it is also important to consider when performing these validations the possibility of compensatory effects among transporters of overlapping specificities (i.e. downregulation or inactivation of a putative drug transporter compensated by upregulation of another with similar substrate specificity), which can obscure the identification of correctly predicted transport mechanisms. Such cases can considerably difficult validations and might require more complex experimental set-ups, such us the generation of cell lines bearing multiple gene knock-outs simultaneously.

A major limitation of drug response prediction is the size and quality of pharmacogenomics datasets. Indeed, while the performance of diverse modeling approaches is generally comparable, input data has been reported as the most important determinant of model performance (Bayer et al, 2013). The generation of new cell line sensitivity data will certainly increase the statistical power of prediction algorithms, but a more important benefit might be obtained from the standardization of laboratory techniques, a more stringent control of experimental conditions and special care in data curation.

A plethora of external variables can affect the quantitative results of cytotoxicity assays, including culture media composition (e.g. serum, glucose concentration, pH), incubation time, seeding cell density, compound dose range, and the specific method used in the assay (Crouch & Slater, 2001; Fellows & O'Donovan, 2007; Weinstein & Lorenzi, 2013). In this sense, some studies have raised concerns about the concordance between cell line sensitivity data coming from different consortia, pointing to severe inconsistencies in these pharmacogenomics datasets and highlighting their negative impact in assessing gene-drug associations using computational methods (Cortés-Ciriano & Bender, 2016; Haibe-Kains et al, 2013; Safikhani et al, 2016). In contrast, others have contested such studies by showing that similar predictors of drug sensitivity can be identified in different datasets if robust statistics and biologically grounded analytical methods are used (Cancer Cell Line Encyclopedia Consortium & Genomics of Drug Sensitivity in Cancer Consortium, 2015; Haverty et al, 2016). Ultimately, these reports suggest that careful attention needs to be paid when comparing and integrating cytotoxicity data obtained in different set-ups in order to derive robust biological conclusions, given the inherent noise and variability of drug sensitivity measurements.

In contrast to drug sensitivity data, molecular profiles of cancer cell lines are highly consistent across datasets (Haibe-Kains et al, 2013). However, given that the most predictive models are based on gene expression, model performance could also be enhanced by a more robust assessment of the transcriptome, for instance, using RNA-seq instead of microarrays to quantify transcript levels. RNA-seq indeed avoids many issues inherent to microarray technology (e.g. cross- and non-specific probe hybridization, high background noise, limited detection range), and is able to detect a larger dynamic range of expression levels, providing absolute instead of relative values, as well as to quantify the whole transcriptome, as it does not rely on a pre-designed set of detection probes (Zhao et al, 2014). Such advantages lead to better estimates of gene expression that have also been shown to be highly reproducible across different laboratories ('t Hoen et al, 2013; Xu et al, 2013). Although all main pharmacogenomics datasets to date provide microarray gene expression data, RNA-seq data for larger numbers of cancer cell lines is becoming available and should therefore be considered in order to improve the prediction power of computational models of drug sensitivity (Garcia-Alonso et al, 2018; Klijn et al, 2015). Nevertheless, probably the most important limitation of using gene expression

for drug sensitivity prediction is the uncertain relationship between mRNA levels and protein abundance, and the relationship of either of the two to actual protein function, especially given the range of transport rates among SLCs and ABCs (Liu et al, 2016; Taslimifar et al, 2017). Quantitative proteomics data of cancer cell lines, which is so far scarce and even harder to obtain for membrane proteins, could at least partially overcome this limitation in the near future (Gholami et al, 2013; Hörmann et al, 2016; Zhao et al, 2017).

Another aspect that can be explored in order to optimize these models is the compound activity parameter used as dependent variable. Most studies reported in the literature use IC_{50} or, as in our case, AUC values, which are generally preferred to the former as they are able to capture both potency and efficacy and seem to be more robust (Fallahi-Sichani et al, 2013). However, these measures are highly dependent on cell division rate, thus varying significantly between experiments due to factors unrelated to the underlying biology, such as differences in growth conditions or assay duration. For this reason, some authors have proposed the use of alternative methods based on the comparison of growth rates that are independent of cell division and would only require small changes in standard experimental protocols, representing a promising alternative to traditional drug response metrics (Hafner et al, 2016).

The second approach we used for interrogating SLC-drug interactions was a CRISPR-Cas9-based genetic screen. The CRISPR-Cas9 system is especially well suited for the generation of pools of cells bearing individual gene knock-outs that can be then tested for a specific phenotype in order to find genomic causative links (Doench, 2018). In our case, we generated a population of cells each lacking a single SLC gene that was subsequently subject to treatment with cytotoxic compounds in order to identify those SLCs that conferred drug resistance, as these could represent new cases of carrier-mediated drug uptake. Moreover, the use of HAP1 cells as a model system, a human haploid cell line that expresses less than half of the total SLC repertoire, can greatly simplify experimental result interpretation, as it avoids confounding allelic effects and reduces, at least partially, the expected substrate redundancy among SLCs (Carette et al, 2011). Furthermore, the compounds used for sensitivity testing were selected with the aim of covering a wide range of physicochemical properties, as shown by the corresponding cheminformatics analysis, and they were tested in multiple concentrations in order to account for dose effects. All these considerations contributed to the design of a

genetic screening approach that enabled the detection of SLC-drug dependencies in a robust and unbiased manner.

In order to obtain a list of resistance-conferring candidate genes, we first assessed enrichment of individual sgRNA guides by applying the DESeq2 method, a common approach for differential gene expression analysis in RNA-seq experiments, and then aggregated guides into genes by using a modified version of the standard gene set enrichment analysis algorithm (GSEA) (Love et al, 2014; Sergushichev, 2016; Subramanian et al, 2005). Although other methods for CRISPR-Cas9 screen analysis exist, this approach has been proven successful in the past (Sedlyarov et al, 2018; Winter et al, 2017) and gave us full control of the different variables affecting our screen. For instance, it enables the consideration of complex designs to model read count variances, including different time points (i.e. before and after drug treatment) and concentrations, thus providing a more accurate estimation of fold changes and probabilities.

The amount of samples retrieved after treatment decreased with increasing drug concentrations, but enrichment was stronger for the higher selective pressures. In total, we could identify ~200 significant SLC-drug associations (FDR <1%), involving 47 compounds at different doses and 101 transporters, and were able to validate the majority of a selected set of high-scoring candidates (19/34) using flow cytometry-based multicolor competition assays (MCA). Some of these associations corresponded to known transport events. For example, the folate transporter SLC19A1 was extremely enriched in all samples treated with antifolates (i.e. methotrexate, pralatrexate, raltitrexed), in agreement with being their main uptake route into cells (Zhao & Goldman, 2013), and treatment with nucleoside analogs (i.e. gemcitabine, cytarabine, decitabine) led to the enrichment of SLC29A1, their main nucleoside transporter expressed in HAP1 cells. Among the newly discovered interactions, a few are likely to consist in direct transport instances, such as the association between the monocarboxylate transporter SLC16A1 and the antimalarials artesunate/dihydroartemisinin or between the glutamine transporter SLC38A5 and the chemotherapy agent cisplatin. Further validation experiments are ongoing in order to prove such cases. In contrast, other associations might rather correspond to SLC-mediated changes in cellular metabolism that affect drug action. As an illustration, the interaction between SLC11A2, an endosomal iron transporter, and the above-mentioned artemisinin derivatives fits the described requirement of iron for the activation

of these compounds (Cui & Su, 2009). Similarly, other cases involved participants of major energetic pathways, such as the glucose transporter SLC2A1 (GLUT1), or the mitochondrial transporters SLC25A3, a phosphate carrier, and MTCH2, a transporter involved in the regulation of apoptosis.

Although this screening approach was successful in identifying multiple transporter-drug interactions, we did not manage to get significant enrichments for a number of compounds. The already mentioned functional redundancy that exists among many SLC transporters, which cannot be resolved in single gene KO set-ups, is a very plausible reason behind this. Multiple gene KO experiments, where several transporters are deactivated simultaneously, might serve to counteract the compensatory effects resulting from such overlapping specificities and uncover the otherwise obscured phenotypes. In this sense, combinatorial CRISPR-Cas9 screens have already been tested, which normally involve the use of dual guide RNAs (Han et al, 2017; Shen et al, 2017; Wong et al, 2016). Alternatively, double KOs could also be obtained by deploying the sgRNA library used in our screen on a previously developed panel of KO cells. However, these combinatorial alternatives also present several drawbacks. On the one hand, the number of combinations increases exponentially, making such screens increasingly more laborious to set up and practically unfeasable for more than triple KOs if large numbers of genes are considered, as it is the case for transporters. On the other hand, the more genes are knocked-out simultaneously in a cell, the more likely this will have a deleterious effect in cell growth that is not related to drug treatment, especially considering the key role of SLCs in mediating the transport of all major biomolecules. In relation to this, another reason for the lack of results for a number of compounds might be that the SLCs involved in their transport or potency are actually essential in HAP1 cells, and thus their KOs are not viable. Using other cell lines as a model could therefore be necessary in order to complement and compare the results obtained in this screen, not restricting them to a specific cellular genotype. Finally, our screen was based on differential cell survival and thus restricted to cytotoxic compounds, but more complex, FACS- or image-based read-outs could also be explored in order to increase the number of drugs and doses tested and uncover other transporter-drug dependencies (Feldman et al, 2018; Parnas et al, 2015).

3.2 Conclusion and future prospects

Altogether, this thesis provides one of the first systematic functional analyses of Solute carriers as a group, focusing on their role as key mediators of drug action. We have addressed our objective by using a double strategy, which consisted, on the one hand, in exploiting by means of computational methods the high information content present in publicly available large-scale pharmacogenomics datasets of panels of molecularly annotated cancer cell lines, and, on the other hand, in applying state-of-the-art techniques for genetic screening and drug sensitivity testing. The combination of both approaches has enabled us to successfully retrieve known SLC-mediated drug transport cases while identifying an important number of previously uncovered relationships, some of which we validated. It is important to bear if mind that, if confirmed, these interactions might not only correspond to transport events but also represent indirect effects by which SLCs, via the transport of endogenous ligands or participation in signaling cascades, are able to modify the cellular inner milieu and metabolism thereby modulating drug response. Either case provides highly valuable information on how transporters affect the mode of action of drug-like compounds.

As we have shown, SLCs constitute an extremely neglected gene group, with a large proportion of their members being barely characterized, yet with an essential role in physiology and an important potential for pharmacology and therapeutics. In this sense, deorphanizing the "SLCome" can highly benefit from systematic analyses, like the one presented here, that consider the whole set of SLCs at once, as these are able to capture functional interdependencies that would be missed with more reductionist approaches. Such strategy has actually proved efficient in the past for other gene families such as GPCRs or kinases (Barr et al, 2009; Edwards et al, 2009). Although we here only focused on the role of SLCs in drug response, integrative studies in the future will have to look systematically into other aspects of SLC biology, including structure, endogenous substrate specificity, participation in signaling pathways, regulation, or association to disease.

In summary, while many of them remain to be further analyzed and validated, the set of SLC-drug associations reported here provides a significant prioritization of potentially pharmacologically relevant relationships, and offers new insights into the mechanism of

drug transport across membranes as well as on the influence of transporters on the pharmacodynamics and pharmacokinetics of therapeutic agents. Such understanding is of key importance in order to achieve more specific and effective treatments in the future, moving towards precision medicine.

References

Aben N, Vis DJ, Michaut M & Wessels LFA (2016) TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. *Bioinformatics* 32: i413-i420

Adli M (2018) The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9: 1911

Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang CZ, Ben-David U, Cook A, Ha G, Harrington WF, Doshi MB, Kost-Alimova M, Gill S, Xu H, Ali LD, Jiang G, Pantel S, Lee Y, Goodale A, Cherniack AD, Oh C et al. (2016) Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* 6: 914-29

Aird EJ, Lovendahl KN, St Martin A, Harris RS & Gordon WR (2018) Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol* 1: 54

Aittoniemi J, Fotinou C, Craig TJ, de Wet H, Proks P & Ashcroft FM (2009) Review. SUR1: a unique ATP-binding cassette protein that functions as an ion channel regulator. *Philos Trans R Soc Lond B Biol Sci* 364: 257-67

Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K & Walter P (2014) Chapter 11: Membrane Transport of Small Molecules and the Electrical Properties of Membranes. In *Molecular Biology of the Cell (Sixth Ed.)* New York: Garland Science

Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106

Andersen J, Stuhr-Hansen N, Zachariassen LG, Koldsø H, Schiøtt B, Strømgaard K & Kristensen AS (2014) Molecular basis for selective serotonin reuptake inhibition by the antidepressant agent fluoxetine (Prozac). *Mol Pharmacol* 85: 703-14

Azuaje F (2017) Computational models for predicting drug responses in cancer research. *Brief Bioinform* 18: 820-829

Baek S, Tsai CA & Chen JJ (2009) Development of biomarker classifiers from high-dimensional data. *Brief Bioinform* 10: 537-46

Bai X, Moraes TF & Reithmeier RAF (2017) Structural biology of solute carrier (SLC) membrane transport proteins. *Mol Membr Biol* 34: 1-32

Bailey DG (2010) Fruit juice inhibition of uptake transport: a new type of food-drug interaction. *Br J Clin Pharmacol* 70: 645-55

Ballatori N, Christian WV, Wheeler SG & Hammond CL (2013) The heteromeric organic solute transporter, OSTα-OSTβ/SLC51: a transporter for steroid-derived molecules. *Mol Aspects Med* 34: 683-92

Barr AJ, Ugochukwu E, Lee WH, King ONF, Filippakopoulos P, Alfano I, Savitsky P, Burgess-Brown NA, Müller S & Knapp S (2009) Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* 136: 352-63

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-12

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603-7

Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, Ebright RY, Stewart ML, Ito D, Wang S, Bracha AL, Liefeld T, Wawer M, Gilbert JC, Wilson AJ, Stransky N, Kryukov GV, Dancik V, Barretina J, Garraway LA et al. (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154: 1151-1161

Bayer I, Groth P & Schneckener S (2013) Prediction errors in learning drug response from gene expression data - influence of labeling, sample size, and machine learning algorithm. *PLoS One* 8: e70294

Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 289–300

Berridge MV, Herst PM & Tan AS (2005) Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol Annu Rev* 11: 127-52

Bhaya D, Davison M & Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45: 273-97

Biemans-Oldehinkel E, Doeven MK & Poolman B (2006) ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett* 580: 1023-35

Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen RL, Ghazal P & Shamu CE (2009) Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* 6: 569-75

Bolotin A, Quinquis B, Sorokin A & Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151: 2551-61

Borst P & Elferink RO (2002) Mammalian ABC transporters in health and disease. *Annu Rev Biochem* 71: 537-92

Bottomly D, Walter NAR, Hunter JE, Darakjian P, Kawane S, Buck KJ, Searles RP, Mooney M, McWeeney SK & Hitzemann R (2011) Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. *PLoS One* 6: e17820

Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W & van Steensel B (2018) Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell* 70: 801-813.e6

Bryan J, Muñoz A, Zhang X, Düfer M, Drews G, Krippeit-Drews P & Aguilar-Bryan L (2007) ABCC8 and ABCC9: ABC transporters that regulate K+ channels. *Pflugers Arch* 453: 703-18

Cancer Cell Line Encyclopedia Consortium & Genomics of Drug Sensitivity in Cancer Consortium (2015) Pharmacogenomic agreement between two cancer cell line data sets. *Nature* 528: 84-7

Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, Kuehne AI, Kranzusch PJ, Griffin AM, Ruthel G, Dal Cin P, Dye JM, Whelan SP, Chandran K & Brummelkamp TR (2011) Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* 477: 340-3

César-Razquin A, Girardi E, Yang M, Brehme M, Saez-Rodriguez J & Superti-Furga G (2018) In silico Prioritization of Transporter-Drug Relationships From Drug Sensitivity Screens. *Front Pharmacol* 9: 1011

César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM & Superti-Furga G (2015) A Call for Systematic Research on Solute Carriers. *Cell* 162: 478-87

Chen L, Li Y, Yu H, Zhang L & Hou T (2012) Computational models for predicting substrates or inhibitors of P-glycoprotein. *Drug Discov Today* 17: 343-51

Chen Z, Shi T, Zhang L, Zhu P, Deng M, Huang C, Hu T, Jiang L & Li J (2016) Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade. *Cancer Lett.* 370: 153–164

Colas C, Ung PMU & Schlessinger A (2016) SLC Transporters: Structure, Function, and Drug Discovery. *Medchemcomm* 7: 1069-1081

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA & Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-23

Cortés-Ciriano I & Bender A (2016) How Consistent are Publicly Reported Cytotoxicity Data? Large-Scale Statistical Analysis of the Concordance of Public Independent Cytotoxicity Measurements. *ChemMedChem* 11: 57-71

Cortes-Ciriano I, Mervin LH & Bender A (2016) Current Trends in Drug Sensitivity Prediction. *Curr Pharm Des* 22: 6918-6927

Cortés-Ciriano I, van Westen GJP, Bouvier G, Nilges M, Overington JP, Bender A & Malliavin TE (2016) Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* 32: 85-95

Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, Bansal M, Ammad-ud-din M, Hintsanen P, Khan SA, Mpindi JP, Kallioniemi O, Honkela A, Aittokallio T, Wennerberg K, NCI DREAM Community, Collins JJ, Gallahan D, Singer D, Saez-Rodriguez J et al. (2014) A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 32: 1202-12

Crouch SP & Slater KJ (2001) High-throughput cytotoxicity screening: hit and miss. *Drug Discov Today* 6: 48-53

Cui L & Su Xz (2009) Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Rev Anti Infect Ther* 7: 999-1013

Dai Z, Sheridan JM, Gearing LJ, Moore DL, Su S, Wormald S, Wilcox S, O'Connor L, Dickins RA, Blewitt ME & Ritchie ME (2014) edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res* 3: 95

Dang Y, Jia G, Choi J, Ma H, Anaya E, Ye C, Shankar P & Wu H (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol* 16: 280

Dean M, Rzhetsky A & Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 11: 1156-66

Deans RM, Morgens DW, Ökesli A, Pillay S, Horlbeck MA, Kampmann M, Gilbert LA, Li A, Mateo R, Smith M, Glenn JS, Carette JE, Khosla C & Bassik MC (2016) Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. *Nat Chem Biol* 12: 361-6

DeGorter MK, Xia CQ, Yang JJ & Kim RB (2012) Drug transporters in drug efficacy and toxicity. *Annu Rev Pharmacol Toxicol* 52: 249-73

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J & Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471: 602-7

Dezi M, Fribourg PF, Di Cicco A, Arnaud O, Marco S, Falson P, Di Pietro A & Lévy D (2010) The multidrug resistance half-transporter ABCG2 is purified as a tetramer upon selective extraction from membranes. *Biochim Biophys Acta* 1798: 2094-101

Di L, Artursson P, Avdeef A, Ecker GF, Faller B, Fischer H, Houston JB, Kansy M, Kerns EH, Krämer SD, Lennernäs H & Sugano K (2012) Evidence-based approach to assess passive diffusion and carrier-mediated drug transport. *Drug Discov Today* 17: 905-12

Di YD, Schafer DW, Cumbie JS & Chang JH (2011) The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq. *Statistical Applications in Genetics and Molecular Biology* 10: 1544-6115

Diaz AA, Qin H, Ramalho-Santos M & Song JS (2015) HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res* 43: e16

Dobson PD & Kell DB (2008) Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule?. *Nat Rev Drug Discov* 7: 205-20

Dobson PD, Patel Y & Kell DB (2009) 'Metabolite-likeness' as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discov Today* 14: 31-40

Doench JG (2018) Am I ready for CRISPR? A user's guide to genetic screens. *Nat Rev Genet* 19: 67-80

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ & Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 32: 1262-7

Dong Z, Zhang N, Li C, Wang H, Fang Y, Wang J & Zheng X (2015) Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* 15: 489

Dooher JE, Schneider BL, Reed JC & Lingappa JR (2007) Host ABCE1 is at plasma membrane HIV assembly sites and its dissociation from Gag is linked to subsequent events of virus production. *Traffic* 8: 195-211

Edwards AM, Bountra C, Kerr DJ & Willson TM (2009) Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol* 5: 436-40

Engelman DM (2005) Membranes are more mosaic than fluid. *Nature* 438: 578-80

Estudante M, Soveral G, Moraisab JG & Benetc LZ (2016) Insights into solute carriers: physiological functions and implications in disease and pharmacokinetics. *Med. Chem. Commun.* 7: 1462-1478

Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW & Sorger PK (2013) Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat Chem Biol* 9: 708-14

Fan F & Wood KV (2007) Bioluminescent assays for high-throughput screening. *Assay Drug Dev Technol* 5: 127-36

Fang Y, Qin Y, Zhang N, Wang J, Wang H & Zheng X (2015) DISIS: prediction of drug response through an iterative sure independence screening. *PLoS One* 10: e0120408

Feldman D, Singh A, Schmid-Burgk J, Mezger A, Garrity A, Carlson R, Zhang F & P B (2018) Pooled optical screens in human cells. *bioRxiv*

Fellows MD & O'Donovan MR (2007) Cytotoxicity in cultured mammalian cells is a function of the method used to estimate it. *Mutagenesis* 22: 275-80

Fets L, Driscoll PC, Grimm F, Jain A, Nunes PM, Gounis M, Doglioni G, Papageorgiou G, Ragan TJ, Campos S, Silva Dos Santos M, MacRae JI, O'Reilly N, Wright AJ, Benes CH, Courtney KD, House D & Anastasiou D (2018) MCT2 mediates concentration-dependent inhibition of glutamine metabolism by MOG. *Nat Chem Biol* 14: 1032-1042

Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, Munoz AG, Pilkington R, Fischer M, Westermann F, Kolch W, Kholodenko BN & Croucher DR (2015) Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci Signal* 8: ra130

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J & Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44: D279-85

Fredriksson R, Nordström KJV, Stephansson O, Hägglund MGA & Schiöth HB (2008) The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. *FEBS Lett* 582: 3811-6

Gadsby DC, Vergani P & Csanády L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* 440: 477-83

Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Ahkmetova L, Pauli A, Montague TG, Zimmerman S, Richter C & Schier AF (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* 9: e98186

Garcia-Alonso L, Iorio F, Matchan A, Fonseca N, Jaaks P, Peat G, Pignatelli M, Falcone F, Benes CH, Dunham I, Bignell G, McDade SS, Garnett MJ & Saez-Rodriguez J (2018) Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. *Cancer Res* 78: 769-780

Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH & Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468: 67-71

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483: 570-5

Gasiunas G, Barrangou R, Horvath P & Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109: E2579-86

Geeleher P, Cox NJ & Huang RS (2014) Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol* 15: R47

Gholami AM, Hahne H, Wu Z, Auer FJ, Meng C, Wilhelm M & Kuster B (2013) Global proteome analysis of the NCI-60 cell line panel. *Cell Rep* 4: 609-20

Giacomini KM & Huang SM (2013) Transporters in drug development and clinical pharmacology. *Clin Pharmacol Ther* 94: 3-9

Gönen M & Margolin AA (2014) Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* 30: i556-63

Goñi FM (2014) The basic structure and dynamics of cell membranes: an update of the Singer-Nicolson model. *Biochim Biophys Acta* 1838: 1467-76

Goodman LS, Brunton LL, Chabner B & Knollmann BC (2011) *Goodman & Gilman's the pharmacological basis of therapeutics (12 Ed.)* New York: McGraw-Hill

Gottesman MM & Pastan I (1988) The multidrug transporter, a double-edged sword. *J Biol Chem* 263: 12163-6

Gray KA, Seal RL, Tweedie S, Wright MW & Bruford EA (2016) A review of the new HGNC gene family resource. *Hum Genomics* 10: 6

Grissa I, Vergnaud G & Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8: 172

Gupta S, Chaudhary K, Kumar R, Gautam A, Nanda JS, Dhanda SK, Brahmachari SK & Raghava GPS (2016) Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: A step towards personalized medicine. *Sci Rep* 6: 23857

Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, Joly JS & Concordet JP (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* 17: 148

Hafner M, Niepel M, Chung M & Sorger PK (2016) Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods* 13: 521-7

Hagenbuch B & Gui C (2008) Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. *Xenobiotica* 38: 778-801

Hagenbuch B & Meier PJ (2004) Organic anion transporting polypeptides of the OATP/ SLC21 family: phylogenetic classification as OATP/ SLCO superfamily, new nomenclature and molecular/functional properties. *Pflugers Arch* 447: 653-65

Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJWL & Quackenbush J (2013) Inconsistency in large pharmacogenomic studies. *Nature* 504: 389-93

Han K, Jeng EE, Hess GT, Morgens DW, Li A & Bassik MC (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol* 35: 463-474

Hardcastle TJ & Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11: 422

Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S & Moffat J (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163: 1515-26

Hart T & Moffat J (2016) BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* 17: 164

Haverty PM, Lin E, Tan J, Yu Y, Lam B, Lianoglou S, Neve RM, Martin S, Settleman J, Yauch RL & Bourgon R (2016) Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* 533: 333-7

Hediger MA, Clémençon B, Burrier RE & Bruford EA (2013) The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med* 34: 95-107

Hediger MA, Romero MF, Peng JB, Rolfs A, Takanaga H & Bruford EA (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *Pflugers Arch* 447: 465-8

Heigwer F, Kerr G & Boutros M (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods* 11: 122-3

Hille F, Richter H, Wong SP, Bratovič M, Ressel S & Charpentier E (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172: 1239-1259

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE & Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33: 510-7

Hirano M, Maeda K, Hayashi H, Kusuhara H & Sugiyama Y (2005) Bile salt export pump (BSEP/ABCB11) can transport a nonbile acid substrate, pravastatin. *J Pharmacol Exp Ther* 314: 876-82

Hodges LM, Markova SM, Chinn LW, Gow JM, Kroetz DL, Klein TE & Altman RB (2011) Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein). *Pharmacogenet Genomics* 21: 152-61

Hoerl AE & Kennard RW (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 12: 55-67

Höglund PJ, Nordström KJV, Schiöth HB & Fredriksson R (2011) The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. *Mol Biol Evol* 28: 1531-41

Hörmann K, Stukalov A, Müller AC, Heinz LX, Superti-Furga G, Colinge J & Bennett KL (2016) A Surface Biotinylation Strategy for Reproducible Plasma Membrane Protein Purification and Tracking of Genetic and Drug-Induced Alterations. *J Proteome Res* 15: 647-58

Horvath P & Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327: 167-70

Hsu PD, Lander ES & Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157: 1262-78

Ibarra KD & Pfeiffer JK (2009) Reduced ribavirin antiviral efficacy via nucleoside transporter-mediated drug resistance. *J Virol* 83: 4538-47

Imaoka T, Kusuhara H, Adachi M, Schuetz JD, Takeuchi K & Sugiyama Y (2007) Functional involvement of multidrug resistance-associated protein 4 (MRP4/ABCC4) in the renal elimination of the antiviral drugs adefovir and tenofovir. *Mol Pharmacol* 71: 619-27

International Transporter Consortium, Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KLR, Chu X, Dahlin A, Evers R, Fischer V, Hillgren KM, Hoffmaster KA, Ishikawa T, Keppler D, Kim RB, Lee CA, Niemi M, Polli JW, Sugiyama Y, Swaan PW et al. (2010) Membrane transporters in drug development. *Nat Rev Drug Discov* 9: 215-36

Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, 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 RK, Liu Q, Mironenko T et al. (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* 166: 740-754

Iqbal M, Audette MC, Petropoulos S, Gibb W & Matthews SG (2012) Placental drug transporters and their role in fetal protection. *Placenta* 33: 137-42

Ishino Y, Shinagawa H, Makino K, Amemura M & Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169: 5429-33

Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, Kafri R, Kirschner MW, Clish CB & Mootha VK (2012) Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* 336: 1040-4

James G, Witten D, Hastie T & Tibshirani R (2013) *An Introduction to Statistical Learning (with Applications in R) (Ed.)* Springer Science+Business Media New York

Jang IS, Neto EC, Guinney J, Friend SH & Margolin AA (2014) Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac Symp Biocomput* 63-74

Jansen R, van Embden JDA, Gaastra W & Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43: 1565-75

Jardetzky O (1966) Simple allosteric model for membrane pumps. *Nature* 211: 969-70

Jiang F & Doudna JA (2017) CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* 46: 505-529

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-21

Johansson U, Sönströd C, Norinder U & Boström H (2011) Trade-off between accuracy and interpretability for predictive in silico modeling. *Future Med Chem* 3: 647-63

Kangas L, Grönroos M & Nieminen AL (1984) Bioluminescence of cellular ATP: a new method for evaluating cytotoxic agents in vitro. *Med Biol* 62: 338-43

Kell DB (2015) What would be the observable consequences if phospholipid bilayer diffusion of drugs into cells is negligible?. *Trends Pharmacol Sci* 36: 15-21

Kell DB (2016) Implications of endogenous roles of transporters for drug discovery: hitchhiking and metabolite-likeness. *Nat Rev Drug Discov* 15: 143

Kell DB, Dobson PD & Oliver SG (2011) Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only. *Drug Discov Today* 16: 704-14

Kell DB & Oliver SG (2014) How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion. *Front Pharmacol* 5: 231

Khalili K, Kaminski R, Gordon J, Cosentino L & Hu W (2015) Genome editing strategies: potential tools for eradicating HIV-1/AIDS. *J Neurovirol* 21: 310-21

Kim Y & Chen J (2018) Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. *Science* 359: 915-919

Kircher M, Sawyer S & Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40: e3

Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, Degenhardt J, Mayba O, Gnad F, Liu J, Pau G, Reeder J, Cao Y, Mukhyala K, Selvaraj SK, Yu M, Zynda GJ, Brauer MJ, Wu TD, Gentleman RC et al. (2015) A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol* 33: 306-12

Koepsell H (2013) The SLC22 family with transporters of organic cations, anions and zwitterions. *Mol Aspects Med* 34: 413-35

Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MDC & Yusa K (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 32: 267-73

Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533: 420-4

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O & Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517: 583-8

König J, Müller F & Fromm MF (2013) Transporters and drug-drug interactions: important determinants of drug disposition and effects. *Pharmacol Rev* 65: 944-66

König R, Chiang Cy, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y & Chanda SK (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 4: 847-9

Korkmaz S, Goksuluk D, Zararsiz G & Karahan S (2017) geneSurv: An interactive web-based tool for survival analysis in genomics research. *Comput Biol Med* 89: 487-496

Kosuri S & Church GM (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* 11: 499-507

Kruh GD & Belinsky MG (2003) The MRP family of drug efflux pumps. *Oncogene* 22: 7537-52

Leslie EM, Deeley RG & Cole SPC (2005) Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol Appl Pharmacol* 204: 216-37

Leuthold S, Hagenbuch B, Mohebbi N, Wagner CA, Meier PJ & Stieger B (2009) Mechanisms of pH-gradient driven transport mediated by organic anion polypeptide transporters. *Am J Physiol Cell Physiol* 296: C570-82

Li W, Köster J, Xu H, Chen CH, Xiao T, Liu JS, Brown M & Liu XS (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* 16: 281

Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M & Liu XS (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15: 554

Lin L, Yee SW, Kim RB & Giacomini KM (2015) SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov* 14: 543-60

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46: 3-26

Liu HC, Goldenberg A, Chen Y, Lun C, Wu W, Bush KT, Balac N, Rodriguez P, Abagyan R & Nigam SK (2016) Molecular Properties of Drugs Interacting with SLC22 Transporters OAT1, OAT3, OCT1, and OCT2: A Machine-Learning Approach. *J Pharmacol Exp Ther* 359: 215-29

Liu Y, Beyer A & Aebersold R (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165: 535-50

Locher KP (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* 23: 487-93

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhim R, Weir BA, Mermel C, Barbie DA, Awad T, Zhou X, Nguyen T, Piqani B, Li C, Golub TR, Meyerson M, Hacohen N et al. (2008) Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A* 105: 20380-5

Mahringer A & Fricker G (2016) ABC transporters at the blood-brain barrier. *Expert Opin Drug Metab Toxicol* 12: 499-508

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE & Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823-6

Mancera-Martínez E, Brito Querido J, Valasek LS, Simonetti A & Hashem Y (2017) ABCE1: A special factor that orchestrates translation at the crossroad between recycling and initiation. *RNA Biol* 14: 1279-1285

Marceau CD, Puschnik AS, Majzoub K, Ooi YS, Brewer SM, Fuchs G, Swaminathan K, Mata MA, Elias JE, Sarnow P & Carette JE (2016) Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* 535: 159-63

Marraffini LA & Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322: 1843-5

Matsson P, Fenu LA, Lundquist P, Wiśniewski JR, Kansy M & Artursson P (2015) Quantifying the impact of transporters on cellular drug permeability. *Trends Pharmacol Sci* 36: 255-62

McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR & Challen GA (2016) Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open* 5: 866-74

Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ & Saez-Rodriguez J (2013) Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* 8: e61318

Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, Goodale A, Lee Y, Ali LD, Jiang G, Lubonja R, Harrington WF, Strickland M, Wu T, Hawes DC, Zhivich VA et al. (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 49: 1779-1784

Mojica FJ, Díez-Villaseñor C, Soria E & Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* 36: 244-6

Mojica FJM, Díez-Villaseñor C, García-Martínez J & Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60: 174-82

Morita Sy & Terada T (2014) Molecular mechanisms for biliary phospholipid and drug efflux mediated by ABCB4 and bile salts. *Biomed Res Int* 2014: 954781

Motohashi H & Inui Ki (2013) Multidrug and toxin extrusion family SLC47: physiological, pharmacokinetic and toxicokinetic importance of MATE1 and MATE2-K. *Mol Aspects Med* 34: 661-8

Neto EC, Jang IS, Friend SH & Margolin AA (2014) The Stream algorithm: computationally efficient ridge-regression via Bayesian model averaging, and applications to pharmacogenomic prediction of cancer cell line sensitivity. *Pac Symp Biocomput* 27-38

Nicolson GL (2014) The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochim Biophys Acta* 1838: 1451-66

Nigam SK (2018) The SLC22 Transporter Family: A Paradigm for the Impact of Drug Transporters on Metabolic Pathways, Signaling, and Disease. *Annu Rev Pharmacol Toxicol* 58: 663-687

O'Hagan S, Wright Muelas M, Day PJ, Lundberg E & Kell DB (2018) GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles. *Cell Syst* 6: 230-244.e1

Okabe M, Szakács G, Reimers MA, Suzuki T, Hall MD, Abe T, Weinstein JN & Gottesman MM (2008) Profiling SLCO and SLC22 genes in the NCI-60 cancer cell lines to identify drug uptake transporters. *Mol Cancer Ther* 7: 3081-91

Ong SH, Li Y, Koike-Yusa H & Yusa K (2017) Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries. *Sci Rep* 7: 7384

Overington JP, Al-Lazikani B & Hopkins AL (2006) How many drug targets are there?. *Nat Rev Drug Discov* 5: 993-6

Palmieri F (2013) The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol Aspects Med* 34: 465-84

Papillon-Cavanagh S, De Jay N, Hachem N, Olsen C, Bontempi G, Aerts HJWL, Quackenbush J & Haibe-Kains B (2013) Comparison and validation of genomic predictors for anticancer drug sensitivity. *J Am Med Inform Assoc* 20: 597-602

Park H, Shimamura T, Miyano S & Imoto S (2014) Robust prediction of anti-cancer drug sensitivity and sensitivity-specific biomarker. *PLoS One* 9: e108990

Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosh I, Sanjana NE, Shalem O, Satija R, Raychowdhury R, Mertins P, Carr SA, Zhang F, Hacohen N & Regev A (2015) A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 162: 675-86

Paull KD, Shoemaker RH, Hodes L, Monks A, Scudiero DA, Rubinstein L, Plowman J & Boyd MR (1989) Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst* 81: 1088-92

Paytubi S, Wang X, Lam YW, Izquierdo L, Hunter MJ, Jan E, Hundal HS & Proud CG (2009) ABC50 promotes translation initiation in mammalian cells. *J Biol Chem* 284: 24061-73

Perland E, Bagchi S, Klaesson A & Fredriksson R (2017) Characteristics of 29 novel atypical solute carriers of major facilitator superfamily type: evolutionary conservation, predicted structure and neuronal co-expression. *Open Biol* 7

Perland E & Fredriksson R (2017) Classification Systems of Secondary Active Transporters. *Trends Pharmacol Sci* 38: 305-315

Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CUT & Pestova TV (2010) The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell* 37: 196-210

Pourcel C, Salvignol G & Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151: 653-63

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP & Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173-83

Qin Y, Chen M, Wang H & Zheng X (2015) A network flow-based method to predict anticancer drug sensitivity. *PLoS One* 10: e0127380

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND & Betel D (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14: R95

Rask-Andersen M, Masuram S, Fredriksson R & Schiöth HB (2013) Solute carriers as drug targets: current use, clinical trials and prospective. *Mol Aspects Med* 34: 702-10

Rask-Andersen M, Masuram S & Schiöth HB (2014) The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu Rev Pharmacol Toxicol* 54: 9-26

Rebsamen M, Pochini L, Stasyk T, de Araújo MEG, Galluccio M, Kandasamy RK, Snijder B, Fauster A, Rudashevskaya EL, Bruckner M, Scorzoni S, Filipek PA, Huber KVM, Bigenzahn JW, Heinz LX, Kraft C, Bennett KL, Indiveri C, Huber LA & Superti-Furga G (2015) SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. *Nature* 519: 477-81

Rees DC, Johnson E & Lewinson O (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* 10: 218-27

Riddick G, Song H, Ahn S, Walling J, Borges-Rivera D, Zhang W & Fine HA (2011) Predicting in vitro drug sensitivity using Random Forests. *Bioinformatics* 27: 220-4

Robey RW, Pluchino KM, Hall MD, Fojo AT, Bates SE & Gottesman MM (2018) Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat Rev Cancer* 18: 452-464

Robey RW, To KKK, Polgar O, Dohse M, Fetsch P, Dean M & Bates SE (2009) ABCG2: a perspective. *Adv Drug Deliv Rev* 61: 3-13

Rubio-Aliaga I & Daniel H (2002) Mammalian peptide transporters as targets for drug delivery. *Trends Pharmacol Sci* 23: 434-40

Safikhani Z, Smirnov P, Freeman M, El-Hachem N, She A, Rene Q, Goldenberg A, Birkbak NJ, Hatzis C, Shi L, Beck AH, Aerts HJWL, Quackenbush J & Haibe-Kains B (2016) Revisiting inconsistency in large pharmacogenomic studies. *F1000Res* 5: 2333

Saier Jr MH, Reddy VS, Tsu BV, Ahmed MS, Li C & Moreno-Hagelsieb G (2016) The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* 44: D372-9

Sanjana NE, Shalem O & Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11: 783-784

Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F & Doench JG (2018) Up, down, and out: optimized libraries for CRISPRa, CRISPRi, and CRISPR-knockout genetic screens. *bioRxiv* 1-24

Sarkadi B, Homolya L, Szakács G & Váradi A (2006) Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. *Physiol Rev* 86: 1179-236

Scheen AJ (2015) Pharmacodynamics, efficacy and safety of sodium-glucose co-transporter type 2 (SGLT2) inhibitors for the treatment of type 2 diabetes mellitus. *Drugs* 75: 33-59

Schlessinger A, Matsson P, Shima JE, Pieper U, Yee SW, Kelly L, Apeltsin L, Stroud RM, Ferrin TE, Giacomini KM & Sali A (2010) Comparison of human solute carriers. *Protein Sci* 19: 412-28

Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, Jones V, Bodycombe NE, Soule CK, Gould J, Alexander B, Li A, Montgomery P, Wawer MJ, Kuru N, Kotz JD, Hon CSY, Munoz B, Liefeld T, Dančík V et al. (2015) Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov* 5: 1210-23

Sebaugh JL (2011) Guidelines for accurate EC50/IC50 estimation. *Pharm Stat* 10: 128-34

Sedlyarov V, Eichner R, Girardi E, Essletzbichler P, Goldmann U, Nunes-Hasler P, Srndic I, Moskovskich A, Heinz LX, Kartnig F, Bigenzahn JW, Rebsamen M, Kovarik P, Demaurex N & Superti-Furga G (2018) The Bicarbonate Transporter SLC4A7 Plays a Key Role in Macrophage Phagosome Acidification. *Cell Host Microbe* 23: 766-774.e5

Sergushichev A (2016) An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*

Seyednasrollah F, Laiho A & Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* 16: 59-70

Shalem O, Sanjana NE & Zhang F (2015) High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16: 299-311

Sharom FJ (2008) ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* 9: 105-27

Sheel A & Xue W (2016) Genomic Amplifications Cause False Positives in CRISPR Screens. *Cancer Discov* 6: 824-6

Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, Sanchez KS, Thomas A, Kuo CC, Du D, Roguev A, Lewis NE, Chang

AN, Kreisberg JF, Krogan N, Qi L et al. (2017) Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods* 14: 573-576

Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB & Vakoc CR (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 33: 661-7

Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, Severinov K, Zhang F & Koonin EV (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* 15: 169-182

Shoemaker RH (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 6: 813-23

Singer SJ & Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175: 720-31

Smith D, Artursson P, Avdeef A, Di L, Ecker GF, Faller B, Houston JB, Kansy M, Kerns EH, Krämer SD, Lennernäs H, van de Waterbeemd H, Sugano K & Testa B (2014) Passive lipoidal diffusion and carrier-mediated cell uptake are both important mechanisms of membrane permeation in drug disposition. *Mol Pharm* 11: 1727-38

Smith DE, Clémençon B & Hediger MA (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol Aspects Med* 34: 323-36

Spahn PN, Bath T, Weiss RJ, Kim J, Esko JD, Lewis NE & Harismendy O (2017) PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci Rep* 7: 15854

Staud F, Cerveny L, Ahmadimoghaddam D & Ceckova M (2013) Multidrug and toxin extrusion proteins (MATE/SLC47); role in pharmacokinetics. *Int J Biochem Cell Biol* 45: 2007-11

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES & Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545-50

Sugano K, Kansy M, Artursson P, Avdeef A, Bendels S, Di L, Ecker GF, Faller B, Fischer H, Gerebtzoff G, Lennernaes H & Senner F (2010) Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov* 9: 597-614

Szakács G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, Bussey KJ, Reinhold W, Guo Y, Kruh GD, Reimers M, Weinstein JN & Gottesman MM (2004) Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. *Cancer Cell* 6: 129-37

Szakács G, Paterson JK, Ludwig JA, Booth-Genthe C & Gottesman MM (2006) Targeting multidrug resistance in cancer. *Nat Rev Drug Discov* 5: 219-34

Tarling EJ, de Aguiar Vallim TQ & Edwards PA (2013) Role of ABC transporters in lipid transport and human disease. *Trends Endocrinol Metab* 24: 342-50

Taslimifar M, Oparija L, Verrey F, Kurtcuoglu V, Olgac U & Makrides V (2017) Quantifying the relative contributions of different solute carriers to aggregate substrate transport. *Sci Rep* 7: 40628

Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58: 267-288

't Hoen PAC, Friedländer MR, Almlöf J, Sammeth M, Pulyakhina I, Anvar SY, Laros JFJ, Buermans HPJ, Karlberg O, Brännvall M, GEUVADIS Consortium, den Dunnen JT, van Ommen GJB, Gut IG, Guigó R, Estivill X, Syvänen AC, Dermitzakis ET & Lappalainen T (2013) Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat Biotechnol* 31: 1015-22

ter Beek J, Guskov A & Slotboom DJ (2014) Structural diversity of ABC transporters. *J Gen Physiol* 143: 419-35

Vasiliou V, Vasiliou K & Nebert DW (2009) Human ATP-binding cassette (ABC) transporter family. *Hum Genomics* 3: 281-90

Vichai V & Kirtikara K (2006) Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat Protoc* 1: 1112-6

Wan Q & Pal R (2014) An ensemble based top performing approach for NCI-DREAM drug sensitivity prediction challenge. *PLoS One* 9: e101183

Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B & Goldenberg A (2014) Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 11: 333-7

Wang L, Li X, Zhang L & Gao Q (2017) Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* 17: 513

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES & Sabatini DM (2015) Identification and characterization of essential genes in the human genome. *Science* 350: 1096-101

Wang T, Wei JJ, Sabatini DM & Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343: 80-4

Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, Chen WW, Lander ES & Sabatini DM (2017) Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* 168: 890-903.e15

Weiner ID & Verlander JW (2010) Molecular physiology of the Rh ammonia transport proteins. *Curr Opin Nephrol Hypertens* 19: 471-7

Weinstein JN, Kohn KW, Grever MR, Viswanadhan VN, Rubinstein LV, Monks AP, Scudiero DA, Welch L, Koutsoukos AD & Chiausa AJ (1992) Neural computing in cancer drug development: predicting mechanism of action. *Science* 258: 447-51

Weinstein JN (2012) Drug discovery: Cell lines battle cancer. *Nature* 483: 544-5

Weinstein JN & Lorenzi PL (2013) Cancer: Discrepancies in drug sensitivity. *Nature* 504: 381-3

Winter GE, Radic B, Mayor-Ruiz C, Blomen VA, Trefzer C, Kandasamy RK, Huber KVM, Gridling M, Chen D, Klampfl T, Kralovics R, Kubicek S, Fernandez-Capetillo O, Brummelkamp TR & Superti-Furga G (2014) The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat Chem Biol* 10: 768-773

Winter J, Breinig M, Heigwer F, Brügemann D, Leible S, Pelz O, Zhan T & Boutros M (2016) caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. *Bioinformatics* 32: 632-4

Winter J, Schwering M, Pelz O, Rauscher B, Zhan T, Heigwer F & Boutros M (2017) CRISPRAnalyzeR: Interactive analysis, annotation and documentation of pooled CRISPR screens. *bioRxiv*

Won S, Morris N, Lu Q & Elston RC (2009) Choosing an optimal method to combine P-values. *Stat Med* 28: 1537-53

Wong ASL, Choi GCG, Cui CH, Pregernig G, Milani P, Adam M, Perli SD, Kazer SW, Gaillard A, Hermann M, Shalek AK, Fraenkel E & Lu TK (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc Natl Acad Sci U S A* 113: 2544-9

Wu A, Xiao T, Fei T, Liu XS & Li W (2018) Reducing False Positives in CRISPR/Cas9 Screens from Copy Number Variations. *bioRxiv*

Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE & Smyth GK (2010) ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* 26: 2176-82

Wu D & Smyth GK (2012) Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* 40: e133

Xu X, Zhang Y, Williams J, Antoniou E, McCombie WR, Wu S, Zhu W, Davidson NO, Denoya P & Li E (2013) Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics* 14 Suppl 9: S1

Yamashita A, Singh SK, Kawate T, Jin Y & Gouaux E (2005) Crystal structure of a bacterial homologue of Na+/Cl–dependent neurotransmitter transporters. *Nature* 437: 215-23

Yan N (2013) Structural investigation of the proton-coupled secondary transporters. *Curr Opin Struct Biol* 23: 483-91

Yan N (2015) Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev Biophys* 44: 257-83

Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, Stratton MR, Benes C, McDermott U & Garnett MJ (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 41: D955-61

Young JD (2016) The SLC28 (CNT) and SLC29 (ENT) nucleoside transporter families: a 30-year collaborative odyssey. *Biochem Soc Trans* 44: 869-76

Yu J, Silva J & Califano A (2016) ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics* 32: 260-7

Yu Z, Surface LE, Park CY, Horlbeck MA, Wyant GA, Abu-Remaileh M, Peterson TR, Sabatini DM, Weissman JS & O'Shea EK (2018) Identification of a transporter complex responsible for the cytosolic entry of nitrogen-containing bisphosphonates. *Elife* 7

Zamek-Gliszczynski MJ, Taub ME, Chothe PP, Chu X, Giacomini KM, Kim RB, Ray AS, Stocker SL, Unadkat JD, Wittwer MB, Xia C, Yee SW, Zhang L, Zhang Y & International Transporter Consortium (2018) Transporters in Drug Development: 2018 ITC Recommendations for Transporters of Emerging Clinical Importance. *Clin Pharmacol Ther*

Zaykin DV (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol* 24: 1836-41

Zhang K, Huentelman MJ, Rao F, Sun EI, Corneveaux JJ, Schork AJ, Wei Z, Waalen J, Miramontes-Gonzalez JP, Hightower CM, Maihofer AX, Mahata M, Pastinen T, Ehret GB, International Consortium for Blood Pressure Genome-Wide Association Studies, Schork NJ, Eskin E, Nievergelt CM, Saier Jr MH & O'Connor DT (2014) Genetic implication of a novel thiamine transporter in human hypertension. *J Am Coll Cardiol* 63: 1542-55

Zhang N, Wang H, Fang Y, Wang J, Zheng X & Liu XS (2015) Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLoS Comput Biol* 11: e1004498

Zhang W, Wu F & Wang B (2018) *MAGeCKFlute: Integrative analysis pipeline for pooled CRISPR functional genetic screens.* R package version 1.2.2.

Zhang Y, Zhang Y, Sun K, Meng Z & Chen L (2018) The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. *J Mol Cell Biol*

Zhao R, Diop-Bove N, Visentin M & Goldman ID (2011) Mechanisms of membrane transport of folates into cells and across epithelia. *Annu Rev Nutr* 31: 177-201

Zhao R & Goldman ID (2013) Folate and thiamine transporters mediated by facilitative carriers (SLC19A1-3 and SLC46A1) and folate receptors. *Mol Aspects Med* 34: 373-85

Zhao S, Fung-Leung WP, Bittner A, Ngo K & Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9: e78644

Zhao W, Li J & Mills GB (2017) Functional proteomic characterization of cancer cell lines. *Oncoscience* 4: 41-42

Zhou F, Zhu L, Wang K & Murray M (2017) Recent advance in the pharmacogenomics of human Solute Carrier Transporters (SLCs) in drug disposition. *Adv Drug Deliv Rev* 116: 21-36

Zhou J, Xu J, Huang Z & Wang M (2015) Transporter-mediated tissue targeting of therapeutic molecules in drug discovery. *Bioorg Med Chem Lett* 25: 993-7

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y & Wei W (2014) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509: 487-91

Zou H & Hastie T (2005) Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67: 301-320