Review

Iterative data multiplexing (IDM) supports elucidation of drug targets from functional genomics screening approaches.

Emma J. Shanks[1*]

[1]Cancer Research UK: Beatson Institute, Garscube Estate, Switchback Road, Bearsden, Glasgow, G61 1BD, UK.

Abstract

The value of conducting high throughput functional genomic screening campaigns has been the subject of some debate over recent years, where a lack of reproducibility and preclusion of the identification of genuine genes of interest under a blanket of non-specific off-target effects have undermined confidence in the technology. High hopes for RNAi-based screening technologies to illuminate a suite of drug targets within different biological scenarios have arguably fallen short of our somewhat unrealistic expectations. But in the age of 'Big Data', where many of us are battling with large and complex datasets, it transpires that where RNAi screening may struggle to work independently, it thrives as a member of a larger team.

Key words:

Iterative data multiplexing, big data, RNAi screening, High Content Analysis, target identification

Introduction

RNA interference (RNAi) is used by many researchers as a tool for reducing the expression of their chosen gene of interest, and quantifying the resulting effect within a defined system. As such, it can provide real insights into key mechanistic pathways mediating effects on or by specific genes. The advent of RNAi screening (including small interfering (siRNA), short hairpin (shRNA) and micro (miRNA)) was professed to be a revolutionary technology which would elucidate a whole array of drug targets in diseases with immense unmet needs. The hope was that these targets would be processed through the drug discovery pipeline and ultimately output as a flood of new therapeutics. But, as with the implementation of any new technology, it takes time to fully understand the system through its application in different contexts, and to appreciate the accompanying perils and pitfalls as it is integrated into current and existing practice.

Resolving the revelation of Off Target Effects.

The realisation that a gene having a quantifiable effect in a screen did not, on further investigation, always correlate well with a reduction in target gene RNA expression, or more worrying still, that the target gene was not expressed in the cell line under investigation, exposed the phenomena of off-target effects (OTEs). OTEs have become a feature of the system which has tarnished the reputation of the technology. As is often the case, once the initial excitement of implementing an innovative technology is superseded by the realisation that it fails to deliver on our (most likely) unrealistic expectations, and we are forced to contemplate why this may be. The resulting unravelling of the origins of the arising issues is, arguably, the most important phase of technology uptake, as it provides real insight into the underlying mechanisms and an advancement of biological knowledge. Much effort has subsequently been invested (largely by companies

[*]**Corresponding author**: Emma Shanks PhD, Head of Screening, The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow, G61 1BD, UK. Email: e.shanks@beatson.gla.ac.uk

Citation: Shanks EJ. Iterative data multiplexing (IDM) supports elucidation of drug targets from functional genomics screening approaches. Cancer Research Frontiers. 2015 Apr; 1(2): 119-126. doi: 10.17980/2015.119

Copyright: @ 2015 Shanks EJ. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The author declares that there is no competing interests.

Received October 13, 2014; Revised February 13, 2015; Accepted March 5, 2015; Published March 27, 2015.

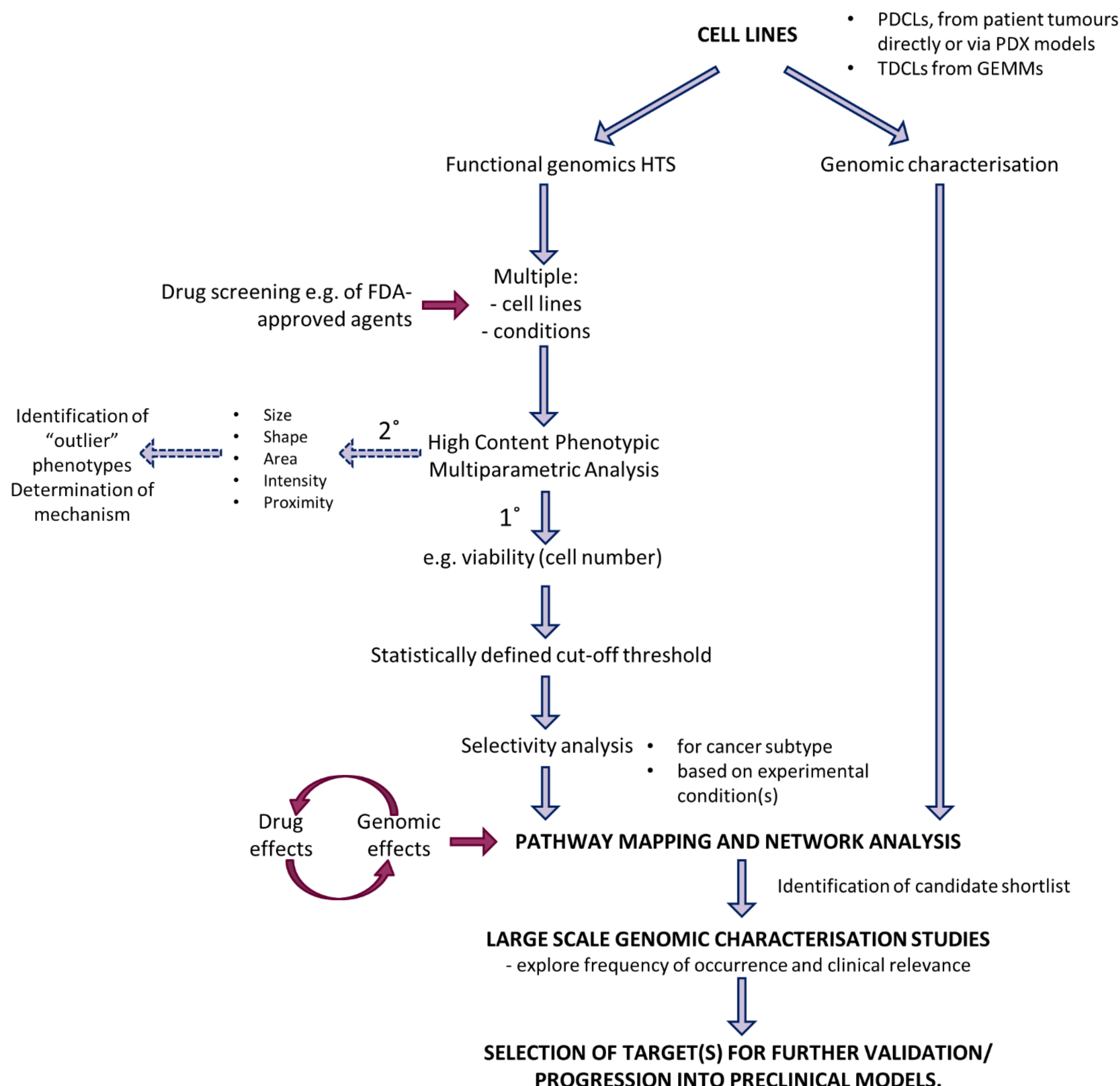

Figure 1: An overview of IDM in the analysis of functional genomics screens.

Multiple sequential and cyclical steps support identification of the most promising preclinical candidates from functional genomic screens. Primary screening data is filtered based on statistically defined rank order, selectivity analysis, pathway mapping and network analysis, and cross referenced with genomic characterisation of cell lines (derived from PDCLs, PDXs, or GEMMs) and additional global analytical techniques. The preclinical relevance of a refined list of putative candidates can then be evaluated against large scale, multi-patient, multi-cancer studies to ensure the progression of the most pertinent targets for subsequent validation studies.

manufacturing large scale RNAi libraries) in improving our understanding of the mechanisms underlying RNAi and OTE (1-4), and the development and incorporation of novel features into the design of reagents for knockdown has had a substantial impact on reducing OTEs. The act of transfecting cells in itself can be considered to be damaging and intrusive, and the resulting cellular response to such an assault can be aggressive, resulting in toxicity. The innate immune component

of the response can identify nucleic acids, and recognise sequences through pattern recognition receptors, including Toll-like receptors (TLR), specifically TLR3 (5, 6) and TLR7-9 (7-9). Those looking to develop RNAi as a therapeutic tool have embraced this dual function with the aim of exploiting its immunity effects. However, when used as an experimental tool, these effects are to be avoided. Chemical modifying the siRNA has proved a somewhat successful strategy in this regard. Modifying the guide strand of the siRNA duplex shifts the dependency for target recognition away from the interaction between the siRNA seed sequence and the target mRNA complex, towards the target-specific region of the oligonucleotide, thereby increasing the 'on target' effect. This approach can also be further extended to the passenger strand (10, 11). The concept of pooling siRNAs is most commonly associated with providing an enhanced knockdown of the target gene by incorporating three or four siRNAs along the length of the corresponding mRNA. However, this approach is also beneficial in reducing OTEs through the simple observation that, due to correlation between the concentration of siRNA used and the likelihood of generating an OTE, a total concentration comprised of lower concentrations of four individual siRNAs rather than a higher concentration of a single siRNA is likely to reduce OTEs. Therefore using pools of siRNAs may not only enhance the efficacy of the target knockdown, but simultaneously reduce putative OTEs.

Understanding the contribution of each component of the pool to the observed phenotype is an important part of evaluating genes of interest for putative OTEs, and manufacturers recommend including this step. Indeed, large scale siRNA libraries are now available in arrayed formats such that three or four individual siRNAs can be tested in a screening assay, thus providing data on multiple individual siRNAs with the same target gene. This approach works extremely well for small, bespoke libraries (e.g. <100 genes). From a large scale screening perspective, this may have been considered too much investment for return in past with regard to screening consumables, data storage and data handling. But the introduction of liquid handling technologies such as acoustic dispensing and higher density microtitre plates, means that throughput becomes less of a limitation, and screening in this format is more accessible. The contribution of each siRNA generating the phenotype of interest is then typically scored. Statistical analyses used in this context include the H score (12) ((number of active siRNAs/total number of siRNAs)*100), Redundant siRNA Analysis (13) (RSA, which assesses the statistical robustness of replicate wells based on rank order of effect), or collective Strictly Standardised Mean Difference (14) (cSSMD). cSSMD is an adaptation of SSMD (15), which assesses the statistical robustness of replicate wells, but importantly, it does this in the absence of a null hypothesis. We employ H score and cSSMD for collective assessment of OTEs and replicate robustness. This is noteworthy because statistics such as RSA (and indeed the commonly implemented Z-score) assume an overall null effect hypothesis, which in many HTS screens may not hold true i.e. in targeted screens, in validation/deconvolution screens. Other computational-based approaches involve evaluating seed sequences for enrichment within all siRNAs screened: again, while useful for large scale i.e. whole genome campaigns, it less suitable for targeted approaches. Similarly, collating a database of identified OTEs has also proved ineffective, as OTEs are often dependent on the biology being evaluated. As such, while modifications to the reagents used have made progress in reducing off-target effects, and tools have been developed to identify sequence enrichments, it is unlikely that this issue will ever be entirely resolved. We must, therefore, reconcile ourselves to fully validate our genes of interest experimentally by testing and overlaying data derived from multiple platforms, reagents and targeted mutants (16). With our expanding working knowledge of the caveats of the system, we are now in an increasingly strengthened position to employ RNAi successfully.

Target identification from large datasets using iterative data multiplexing (IDM).

These days, when we screen, we screen big. Our libraries encompass millions of small molecules, developed drugs, natural products and functional genomic approaches, and we often conduct *in vitro* screening campaigns using several cell types, under multiple conditions and with combinations of agents, which multiplies the number of outputs. Our screening assays have become increasingly data-rich too. The embracing of High Content Imaging (HCI) and Analysis (HCA), albeit slow to encompass the full extent of the capability (17) has resulted in massively detailed phenotypic characterisation of our cells of choice, where a suite of parameters can be quantified from a simple single stain, and moreover multiple markers and indicators can be used to provide detailed insight into the underlying the mechanisms in question. With the implementation of low attachment plates and spheroid technologies, we are no longer restricted to a somewhat artificial 2D

platform, thus further escalating the data volume. We are also not constrained within the limits of our own datasets. There is now a wealth of publically accessible data available, pertaining to genomic characterisation and its correlation with patient prognosis: this is currently most evident when engaging in cancer research. The age of "Big Data" has really infiltrated the screening community in an unprecedented and somewhat unexpected way. So how do we identify the most apposite targets from such a volume of data?

When conducting any type of high throughput screen there is always a primary output, and in target-discovery based screening approaches for cancer, this often constitutes a measure of viability, for example, cell number. The inclination with any large data set is to create a list based on rank order, and set an arbitrary or statistically defined threshold based on the overall dataset. While there may be some genes which are extremely effective in reducing cell number, they may not present the best targets for further study from both a mechanistic or putative therapeutic perspective, as they frequently encompass components essential to the viability of all cells. There are a number of structured approaches which can be iteratively applied to such datasets to define a shortlist of genes, and these are summarised in Figure 1. Three complementary approaches demonstrate utility in deriving a preliminary target 'hit' list:

i) Filtering based on selectivity for the disease model of interest, for example, by cross referencing effects of knocking down candidate genes in other cell lines /disease types studied. This can be achieved through exploration of the literature, or through database mining either using internal databases if screens are conducted within a dedicated facility or unit, or through mining external databases, such as GenomeRNAi (18).

ii) Filtering based on internal conditions included within the screen. RNAi approaches are (arguably) at their best when one asks a specific question of them. For example, instead of asking 'knockdown of which genes kill cell line X', we may ask 'knockdown of which genes kill cell line X under low oxygen conditions in the presence of drug Y?'. This could include, for example, screening in the presence or absence of an agent of interest (such as a known drug), screening in a drug sensitive cell line compared with a drug insensitive/resistant cell line, or screening under different environmental conditions (normal oxygen compared with low oxygen, with radiation treatment compared without radiation exposure).

iii) Implementing pathway analysis software such as Metacore (Thomson Reuters) or Ingenuity Pathway Analysis (IPA) to identify effective genes which may indicate a known pathway is affected. This approach can be extremely useful for identifying genes associated with a specific pathway which provides information as to whether a cell line may be dependent on a core hub or an extended network, rather than discrete genes, and may work well using a subset of genes defined from i) and ii).

Further to this, screening gene libraries in parallel with drug libraries may offer a cross-platform validation approach, notably when screening drugs of libraries where the target(s) are well established, such as FDA-approved collections. Here, functional genomics approaches can be used to assess the effect of reducing the expression of the key target(s) associated with the drug and thus validating that the drug may be acting via the predicted mechanism of action. Reciprocally, a drug which is known to inhibit a gene candidate of interest may prove useful as a tool agent for exploring the underlying mechanism, whilst holding potential for development as a therapeutic candidate.

In addition to the aforementioned, inclusion of additional information on the cell line(s) themselves is a powerful approach to guiding target selection. Conducting genomic characterisation of cell lines is becoming increasingly more commonplace, and arms us with a much broader understanding of the genetic backdrop within which we are screening. This provides great insight into the relevance of identified genes and pathways of interest, even more so when any screening conditions are replicated in the genomic analysis i.e. in the presence/absence of a drug or environmental condition. Here, we can apply our hypothesis across two independent data platforms and interrogate datasets for genes which are relevant in both contexts, for example, genes which support viability *and* which are upregulated under low oxygen conditions may present as promising targets. This approach is, of course, not limited to genomic characterisation, and the use of global analytical techniques to profile essential cellular processes are also hugely informative (19). The iterative use of pathway analysis also finds utility here, whereby it can support identification of affected pathways in this wider context.

Further to identifying putative targets/pathways which may be relevant in the cell lines of study, scientists also now have access to a wealth of data to support determination of the clinical relevance. Cross-referencing candidates with literary and

microarray databases (20) remain valuable and informative approaches, but in the present day, targets identified using an IDM approach can subsequently be cross referenced with large scale genomic characterisation studies, such as those published by The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC) and the Catalogue of Somatic mutations in Cancer (COSMIC), and excellent data visualisation tools, such as cBioPortal and Oncomine. These projects have supported comprehensive genomic characterisation of thousands of patient samples across multiple cancer types, with subsequent analysis of the frequency and clinical impact of the observed alterations. As such, they provide an invaluable resource to support prioritisation of a refined list of putative candidates. Equally, these datasets can themselves be a starting point from which global bioinformatics analysis can be conducted, with functional genomic approaches providing the biological validation (21, 22). Collectively, use of IDM in the context described supports translation of multi-platform global datasets towards clinically relevant therapeutics for patient benefit.

RNAi screening approaches to target identification in predictive pre-clinical disease models.

Use of cell lines in the laboratory environment has supported fundamental discoveries of pathways and mechanisms. However, when the aim is to identify innovative therapeutics, tractability from *in vitro* efficacy to clinical efficacy has a disappointing success rate. Furthermore, the majority of clinical failures occur towards the latter Phase trials with between 5% (2004, (23)) and 13.4 % (2013, (24)) of cancer drugs being approved to registration. A refocusing on the investment in preclinical validation stages (25) and improving predictivity of preclinical and pharmacological models has attempted to address this shortfall, and in this scenario, it may prove more fruitful to use models using tumor mass taken directly from the patient and transplanting them into immunocompromised mice followed by serial passage (patient derived xenografts (PDXs)), and/or generate cell lines to support *in* vitro experimentation (Patient derived cell lines (PDCLs)). Both examples have advantages within their remit. PDXs are gaining credence as a superior model for predicting pharmacological effects in patients (26, 27), and are attractive because they can recapitulate essential components of both the tumor, whereby they encompassing heterogeneity and intra-tumoral hierarchy, and interactions with the stromal microenvironment. However, they cannot exclude the effects of interactions between the murine host and the human sample, and offer little information regarding the immune response to certain agents. There is also debate as to the clinical relevance of doses which can be administered to mice. The use of Genetically Engineered Mouse Models (GEMMs), whereby use of transgenics results in the spontaneous development of tumors, goes some way to addressing some of the issues arising in PDXs, namely around the role of the immune system in testing therapeutic candidates. However, GEMMs are reliant on tumors arising due to the alteration in one or two key driver oncogenes, and it is not yet fully understood how representative of the heterogeneity and pathology observed in patient tumors these are (28). As such, combining these two preclinical approaches may provide a complementary platform to support therapeutic validation (29).

Cell lines derived from patient tumors (PDCLs, directly from patient samples or from PDX models), and from GEMMs offer an extension to these preclinical models which can support the integration of *in vitro* functional genomics-based target discovery platforms into this pipeline, for example, in identifying synthetically lethal candidates, and in improving the efficacy of existing therapeutics, as lack of sensitivity, whether innate or acquired, is a fundamental failing of cancer treatments. Traditionally, *in vitro* RNAi screening approaches constitute the initiating point for such endeavours, but increasingly they are finding their place alongside mouse models. Here, use of RNAi techniques can support target discovery within a more clinically tractable model, underscored by genomic and proteomic characterisation: collectively these approaches are useful for identifying targets in, for example, drug sensitive vs drug resistant tumors, and moreover for the potential identification of tumor biomarkers. Moreover shRNA approaches have incorporated genetic knockout screening into *in vivo* models, both through transplantation of transfected cell lines into mice (30) and pooled shRNA libraries directly into murine tumors (31), bridging a long-recognised gap between *in vitro* and *in vivo* techniques. Use of genetic knockdown/out tools can also constitute an iterative process, whereby target identification approaches can be used if a patient develops resistance or presents with disease recurrence. In an age of cheaper and more sensitive genomic sequencing (32), initiatives are underway to embrace these approaches across large scale patient numbers, with the overarching aim of identifying signatures which will identify the most apposite

treatment to give that patient as a first line therapy (33-35). Collective use of these preclinical tools will provide a much broader comprehensive understanding of the cancer genome through genomic characterisation, while sequencing of hundreds and thousands of different tumors across multiple cancer types builds our global understanding of tumor biology and adaptations made when challenged with therapeutic agents which may underlie drug resistance.

Conclusion

Translating functional genomics screening campaigns towards clinically relevant therapeutics has proved extremely challenging. The discovery of off-target effects, and the extent to which they can preclude the identification of genuine targets of interest, has led to considerable effort to appreciate the underlying biology of this phenomena, while refining reagents to reduce the prevalence of such effects. While considerable progress has been made in this regard, it is unlikely that the occurrence of OTEs will be fully resolved, and stratified deconvolution of putative genes of interest will always remain an essential component of hit validation.

Identifying the most apposite targets in the context of the experimental system presents with further challenges, both through the volume of data generated from a large scale functional genomics screen, coupled with the multiple parameters which are increasingly implemented to quantify resulting cellular phenotypes. Cross-referencing with other datasets and platforms can provide a useful tool to refine lists based on selectivity for the biology or experimental conditions being studied, and this can be approached using data generated in-house, or by cross-referencing to databases collating reported phenotypic effects. The application of pathway and network mapping software is particularly useful on these preliminary 'hit' lists, where enrichment for genes within a known pathway may be evident, indicating it as a biologically-relevant avenue for further study.

Genomic characterisation of the cell lines being evaluated provides an excellent background within which to assess the relevance of genes within a refined list, especially when genomic characterisation is carried out in the presence of the experimental conditions included within the screen itself, for example oxygen level or exposure to radiation. Moreover, large efforts to conduct broad scale genomic analyses across thousands of patients with multiple cancer types has facilitated a mechanism by which identified gene candidates can be assessed for putative clinical relevance. This is a truly outstanding resource, and provides much needed insight as to which targets to progress from hit discovery into hit validation, based on those which will benefit a majority, or even a stratified group of cancer patients.

Collectively, the strategies described provide an IDM toolkit to refine lists of effective genes from functional genomics screening datasets and other large scale datasets, and provide essential information as to which genes may be the most pertinent to prioritise for further study. As such, RNAi screening approaches find their place as a player on a carefully strategized team, which provides collaborative insight into cancer biology and cancer target discovery.

Abbreviations

RNAi	RNA interference;
IDM	iterative data multiplexing;
siRNA	small interfering RNA;
shRNA	short hairpin RNA;
miRNA	microRNA;
OTE	off-target effect;
PDCL	patient derived cell line;
TDCL	tumor derived cell line;
GEMM	genetically engineered mouse model;
HCA	High Content Analysis;
BrdU	5-Bromo-2-DeoxyUridine;
PARP	poly ADP ribose polymerase;
IPA	Ingenuity Pathway Analysis.

Acknowledgements.

The author wishes to acknowledge Dr. Lynn McGarry, Dr Kay Hewit, Daniel James and Matthew Davidson in the CRUK Beatson Institute Screening Facility. Additional thanks to Michael Howell (London Research Institute) for discussions around the uptake and implementation of RNAi screening over time.

References

1. Sigoillot FD, King RW. Vigilance and validation: Keys to success in RNAi screening. ACS Chem Biol. 2011 Jan 21;6(1):47-60. DOI: 10.1021/cb100358f.

2. Anderson E, Boese Q, Khvorova A, Karpilow J. Identifying siRNA-Induced Off-Targets by Microarray Analysis. In: Barik S, editor. RNAi. Methods in Molecular Biology™. 442: Humana Press; 2008. p. 45-63.

3. Qiu S, Adema CM, Lane T. A computational study of off-target effects of RNA interference. Nucleic Acids Res. 2005 Mar 30;33(6):1834-47. DOI: 10.1093/nar/gki324.

4. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y, et al. 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat Methods. 2006 Mar;3(3):199-204. DOI: 10.1038/nmeth854.

5. Choe J, Kelker MS, Wilson IA. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science. 2005 Jul 22;309(5734):581-5. DOI: 10.1126/science.1115253.

6. Bell JK, Askins J, Hall PR, Davies DR, Segal DM. The dsRNA binding site of human Toll-like receptor 3. Proc Natl Acad Sci U S A. 2006 Jun 6;103(23):8792-7. DOI: 10.1073/pnas.0603245103.

7. Dalpke A, Helm M. RNA mediated Toll-like receptor stimulation in health and disease. RNA Biol. 2012 Jun;9(6):828-42. DOI: 10.4161/rna.20206.

8. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. Science. 2004 Mar 5;303(5663):1529-31. DOI: 10.1126/science.1093616.

9. Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, et al. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. Science. 2004 Mar 5;303(5663):1526-9. DOI: 10.1126/science.1093620.

10. Chen PY, Weinmann L, Gaidatzis D, Pei Y, Zavolan M, Tuschl T, et al. Strand-specific 5'-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. RNA. 2008 Feb;14(2):263-74. DOI: 10.1261/rna.789808.

11. Bramsen JB, Laursen MB, Nielsen AF, Hansen TB, Bus C, Langkjaer N, et al. A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. Nucleic Acids Res. 2009 May;37(9):2867-81. DOI: 10.1093/nar/gkp106.

12. Bhinder B, Djaballah, H. A Simple Method for Analyzing Actives in Random RNAi Screens: Introducing the “H Score” for Hit Nomination & Gene Prioritization. Combinatorial Chemistry & High Throughput Screening. 2012;15(9):686-704.

13. Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, et al. A probability-based approach for the analysis of large-scale RNAi screens. Nat Methods. 2007 Oct;4(10):847-9. DOI: 10.1038/nmeth1089.

14. Zhang XD, Santini F, Lacson R, Marine SD, Wu Q, Benetti L, et al. cSSMD: assessing collective activity for addressing off-target effects in genome-scale RNA interference screens. Bioinformatics. 2011 Oct 15;27(20):2775-81. DOI: 10.1093/bioinformatics/btr474.

15. Zhang XD. Strictly Standardized Mean Difference, Standardized Mean Difference and Classicalt-test for the Comparison of Two Groups. Statistics in Biopharmaceutical Research. 2010;2(2):292-9. DOI: 10.1198/sbr.2009.0074.

16. Buehler E, Chen YC, Martin S. C911: A bench-level control for sequence specific siRNA off-target effects. PLoS ONE. 2012;7(12):e51942. DOI: 10.1371/journal.pone.0051942.

17. Singh S, Carpenter AE, Genovesio A. Increasing the Content of High-Content Screening: An Overview. J Biomol Screen. 2014 Apr 7;19(5):640-50. DOI: 10.1177/1087057114528537.

18. Schmidt EE, Pelz O, Buhlmann S, Kerr G, Horn T, Boutros M. GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. Nucleic Acids Res. 2013 Jan;41(Database issue):D1021-6. DOI: 10.1093/nar/gks1170.

19. Schug ZT, Peck B, Jones DT, Zhang Q, Grosskurth S, Alam IS, et al. Acetyl-CoA synthetase 2 promotes acetate utilization and maintains cancer cell growth under metabolic stress. Cancer Cell. 2015 Jan 12;27(1):57-71. DOI: 10.1016/j.ccell.2014.12.002.

20. Yang Y, Adelstein SJ, Kassis AI. Target discovery from data mining approaches. Drug Discov Today. 2009 Feb;14(3-4):147-54. DOI: 10.1016/j.drudis.2008.12.005.

21. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014 Aug 28;158(5):1199-209. DOI: 10.1016/j.cell.2014.07.027.

22. Jerby L, Ruppin E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. Clin Cancer Res. 2012 Oct 15;18(20):5572-84. DOI: 10.1158/1078-0432.CCR-12-1856.

23. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004;3(8):711-6.

24. DiMasi JA, Reichert JM, Feldman L, Malins A. Clinical approval success rates for investigational cancer drugs. Clin Pharmacol Ther. 2013 Sep;94(3):329-35. DOI: 10.1038/clpt.2013.117.

25. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol.

2011 Mar;162(6):1239-49. DOI: 10.1111/j.1476-5381.2010.01127.x.

26. Fiebig HH VV, Korrat A, Foucault F, Beckers T. Predictive gene signatures for bevacizumab and cetuximab as well as cytotoxic agents. Int J Clin Pharmacol Ther. 2012;50(1):70-1. DOI: 10.5414/CPP50070.
27. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med. 2004 Jul 22;351(4):337-45. DOI: 10.1056/NEJMoa033025.
28. Olive KP, Tuveson DA. The use of targeted mouse models for preclinical testing of novel cancer therapeutics. Clin Cancer Res. 2006 Sep 15;12(18):5277-87. DOI: 10.1158/1078-0432.CCR-06-0436.
29. Malaney P, Nicosia SV, Dave V. One mouse, one patient paradigm: New avatars of personalized cancer therapy. Cancer Lett. 2014 Mar 1;344(1):1-12. DOI: 10.1016/j.canlet.2013.10.010.
30. Gao H, Chakraborty G, Lee-Lim AP, Mavrakis KJ, Wendel HG, Giancotti FG. Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. Proc Natl Acad Sci U S A. 2014 Nov 18;111(46):16532-7. DOI: 10.1073/pnas.1403234111.
31. Rudalska R, Dauch D, Longerich T, McJunkin K, Wuestefeld T, Kang TW, et al. In vivo RNAi screening identifies a mechanism of sorafenib resistance in liver cancer. Nat Med. 2014 Oct;20(10):1138-46. DOI: 10.1038/nm.3679.
32. Bennett ST, Barnes C, Cox A, Davies L, Brown C. Toward the 1,000 dollars human genome. Pharmacogenomics. 2005 Jun;6(4):373-82. DOI: 10.1517/14622416.6.4.373.
33. Graham JS, Jamieson NB, Rulach R, Grimmond SM, Chang DK, Biankin AV. Pancreatic cancer genomics: where can the science take us? Clin Genet. 2014 Nov 12:12536. DOI: 10.1111/cge.12536.
34. Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer. Nat Rev Clin Oncol. 2011 Mar;8(3):142-50. DOI: 10.1038/nrclinonc.2010.220.
35. International Cancer Genome C, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. Nature. 2010 Apr 15;464(7291):993-8. DOI: 10.1038/nature08987.