A decade of RNAi screening: too much hay and very few needles

RNAi screening is arguably the fastest growing field with the premise to better understand gene function at the genome level. It has been hailed as the second genomics wave, and in combination with the human genome-sequencing projects, would constitute the holy grail of modern genetics.

By Bhavneet Bhinder and Dr Hakim Djaballah

Despite more than a decade of trials and tribulations, RNAi screening today finds itself surrounded more by controversy questioning breakthrough discoveries of gene(s) allegedly identified through this random process than success and celebration, at a time where there is a critical need for novel gene targets to fight disease.

A noble technology enters the race to fight disease

In 1995, Guo and Kemphues successfully implemented an anti-sense RNA-based knockdown of par-1 in *C elegans*[1], perhaps the lining foundations towards revolutionising functional genomics research. Three years later, Fire and Mello specifically reported on double-stranded RNA molecules triggering suppression of gene expression at the mRNA level; a process later coined RNA interference (RNAi); a riveting discovery for which Fire and Mello were awarded the Nobel Prize in Physiology or Medicine in 2006 for the potential RNAi held[2]. The discovery of RNAi opened the door to functional genomics screens to elucidate novel gene functions and pathways at an unprecedented rate since 2000. RNAi and a vast new world of tiny regulatory RNA molecules have profoundly changed the way we think about gene regulation. It is also worth noting here that since the early stage of development of RNAi screening technologies, the occurrence of unintended off-target effect (OTE) has been pointed out as a problem that causes inherent rates of false positives[3,4], but has been largely ignored.

In 2001, Mattick referred to the non-coding RNAs (ncRNAs) as the architects of the eukaryotic complexity, and indeed so since 98% of the transcriptome is constituted of ncRNAs acting as modulators of gene expression[5]. To date, at least 2,208 ncRNA families are known[6], most significant of the families are ~22 nucleotide (nt) long microRNAs (miRNAs), which target ~60% of the genome[7]. The ncRNAs are drawing attention not only in terms of maintenance of biological functions but have also been shown to play a role in repertoire of diseases including cancer[5,8]. These

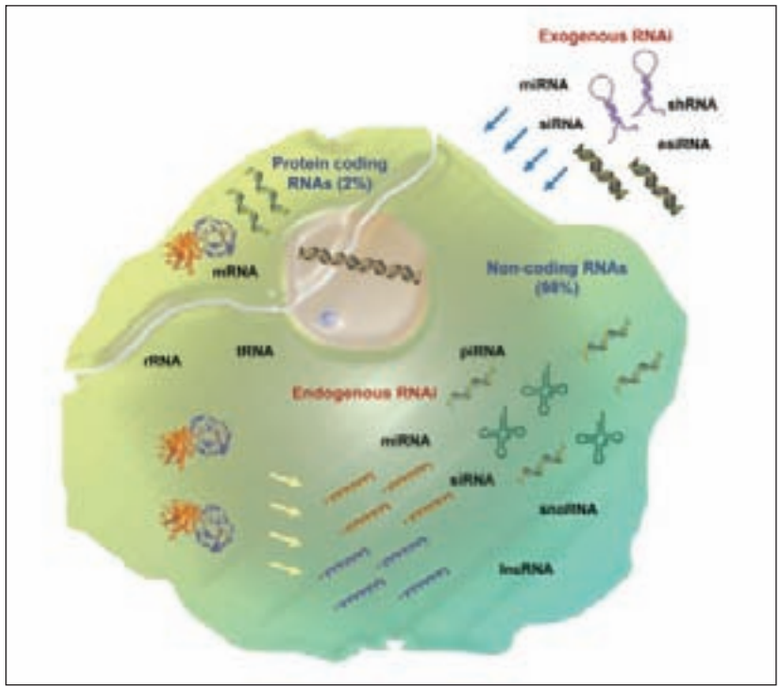

Figure 1: Illustration of an intracellular biological system in a state of equilibrium and its interplay with exogenous RNAi

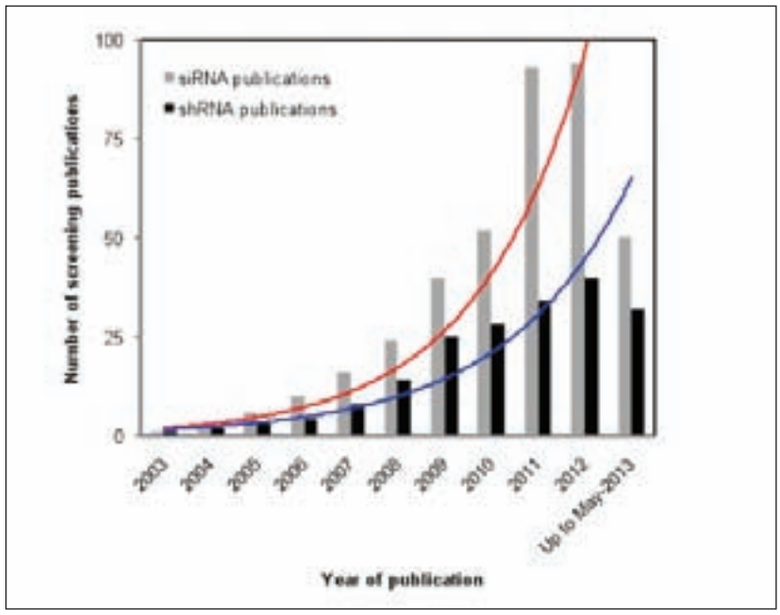

Figure 2: RNAi publications over the past decade. The number of publications has grown on average two-fold every year and yielded a plethora of gene targets. The question is whether this trend yielded any novel therapeutics

observations indicate a complex endogenous RNAi in action that tightly orchestrates cell fate. On the other hand, even though all efforts have been concentrated on studying the residual 1-2% of the proteome, the exact number of genes in the genome remains elusive[9,10]. In addition, constant updates in gene records pose further challenges in defining the human genome set targeted and studied extensively to elucidate gene function; bringing into question the target versus the targeted. Progress in RNAi technologies has opened avenues to elicit gene silencing via introduction of exogenous small interfering RNA (siRNA) duplex, plasmid-vector-based short hairpin RNA (shRNA) hairpins, miRNA mimics or endoribonuclease-prepared siRNA (esiRNA) duplexes (**Figure 1**). Exogenous RNAi duplexes, based on their sequence homology with the intracellular mRNA molecules, are believed to enable silencing of specific genes amidst this tightly regulated intracellular system. This begs the question as to whether a level of selectivity in knocking down a target gene can be achieved without altering the existing intracellular equilibrium. Furthermore, is the observed perturbation specific or merely an outcome of synergy and non-specific knockdowns within the cell; thus, totally independent of the intended gene target? These questions remain unanswered.

Traditionally, high throughput screening methodologies have been associated with screening chemicals, and have evolved over a course of several decades through advances in synthetic chemistry, high density plate formats, instrumentation, automation and informatics tools to perform, acquire and manage screening data output. In a relatively short period of time, the RNAi screening industrialisation phase was complete as it borrowed nearly all concepts from the well-established chemical screening operations without a solid validation as to RNAi screening input and data output in terms of both quality assurance as to the assay, and quality control as to the screen performance, raw data analysis and ultimately 'active gene' nomination. This has resulted in many RNAi screens in mammalian cells conducted in various fields to study various biological processes, such as stem cell biology, signal transductions, RNA metabolism, immunity, host-pathogen interactions, predictive biomarker and responder population of both existing and novel drugs, and so on[11-17]. Based on a very recent survey, ~580 publications have been reported in the literature to date (**Figure 2**). But the novel gene targets, if any, that these RNAi screens discovered should outweigh all the efforts and investments made in the technology thus far.

A decade long of trials and tribulations: any targets at the finish line?

The classical methods of studying gene function were limited to mutagenesis and knockout animal models, homologous recombination, anti-sense DNA technologies and microarrays; all of which were time consuming, labour-intensive and cost prohibitive[18,19]. The introduction of the RNAi technology quickly surpassed these traditional practices enabling investigators to undertake genome-wide systematic and simultaneous gene knockdowns within much shorter timeframes resulting in a record of ~580 publications reporting a plethora of promising therapeutic gene targets (**Figure 2**). The vast majority of them used either siRNA duplexes or shRNA hairpins and vary with regards to library coverage from focused up to genome scale. As for the screening formats, both arrayed/singles as well as pooled formats were employed using a repertoire of assay readouts, and with a prominence of loss-of-function screens. Equally intriguing is the variety of hit nomination data analysis methodologies applied with up to 33 different methods used[20]. With such a gigantic undertaking to help fight disease through identifying novel gene targets for drug discovery; the dreaded question of how many novel and useful therapeutic targets have been obtained from this random screening and entered the discovery process is worth asking. The unfortunate answer is simply none thus far (**Table 1**). On reviewing published reports, it became obvious that on average an RNAi screen would yield anywhere from hundreds to thousands of gene hits[21-31]; understandably, a painstaking task to attempt to validate each one of them. Ironically, what is emerging is an intriguing 'funnel' which magically narrows down the gene list from a large number to a single gene target that is deemed high value, validated and ready for prime time therapeutic discovery. Among these many gene targets reported in high impact journals, only a handful were further studied, few clearly refuted and none have been translated into therapeutic interventions as the final outcome (**Table 1**).

STK33, one of these high value gene targets published in 2009, fell prey to validation scrutiny due to its failure to confirm in independent validations by others[30,32-34]. It was identified as a promising target for KRAS-dependent cancers by Scholl from an arrayed shRNA hairpin screen conducted using a focused TRC library in NOMO-1 cells[30]. Noticeably, it was totally absent from the list of published hits corresponding to the screens performed by others[29,31], who happen to be colleagues from the same affiliation and using the same RNAi library. In 2011, a group from Amgen lead by Babij refuted the claims made by Scholl and concluded that STK33 is not an essential gene target in KRAS dependent cancers[33]. In an effort to restore the original claims and protect the technology, Luo, from the same affiliation as Scholl, attempted to validate the role of STK33 in two KRAS-mutated cell lines (NOMO-1 and SKM-1) by inhibiting its activity using specific kinase inhibitors; surprisingly, their results led to no definitive conclusions with regards to the STK33 involvement in KRAS-dependent cancers[34], and was well in line with the Amgen's group conclusions[33]. TBK1 was also reported by Barbie and coworkers as another high value gene target for KRAS-dependent cancers resulting from an arrayed shRNA screen using the same library and performed against 19 cell lines including NOMO-1 cells[31]. Using a similar approach, Scholl from the same group also published its screening results using the same NOMO-1 cell line, but TBK1 was surprisingly absent[30]. Additionally, another discrepancy was found in genome-wide screens reported by Cheung from the same group using the

Table 1: List of representative gene targets identified from random RNAi screening with no therapeutic outcome

RNAi Technology	Screening Format	Gene Target	Therapeutic Area	Further Validation	Outcome
shRNA	Arrayed	GSK-3α[71]	Oncology	Unknown	None
		RPS14[72]	**Oncology**	**Refuted[74]**	**None**
		STK33[30]	**Oncology**	**Refuted[33,34]**	**None**
		TBK1[30]	**Oncology**	**Refuted[30]**	**None**
	Pooled	BRD4[66,75]	Oncology	Unknown	None
		CARD11[76]	Oncology	Yes[67]	None
		GAS1[75]	Oncology	Unknown	None
		IGFBP7[77]	**Oncology**	**Refuted[78,79]**	**None**
		IKBKE[80]	Oncology	Unknown	None
		NF1[81]	Oncology	Unknown	None
		PAX8[82]	Oncology	Unknown	None
		PITX1[83]	Oncology	Unknown	None
		PLK1[27]	Oncology	Unknown	None
		RES1[84]	Oncology	Unknown	None
		USP15[85]	Oncology	Unknown	None
		WEE1[86]	Oncology	Unknown	None
		WT1[87]	Oncology	Unknown	None
siRNA	Singles	BAP31[67]	Anti-viral	Unknown	None
		EIF[67]	Anti-viral	Unknown	None
	Pooled	**CDLL1[88]**	**Anti-viral**	**Refuted[88]**	**None**
		CDL4A3BP[89]	Oncology	Unknown	None
		EGFR[67]	Anti-viral	Unknown	None
		EphA2[67]	Anti-viral	Unknown	None
		IFITM3[90]	Anti-viral	Yes[90]	None
		LMTK3[91]	Oncology	Yes[91,92]	None
		NAMPT[93]	**Oncology**	**Refuted[95]**	**None**
		SCD1[96]	**Oncology**	**Refuted[97]**	**None**
		TPX2[98]	Oncology	Yes[98, 99]	None

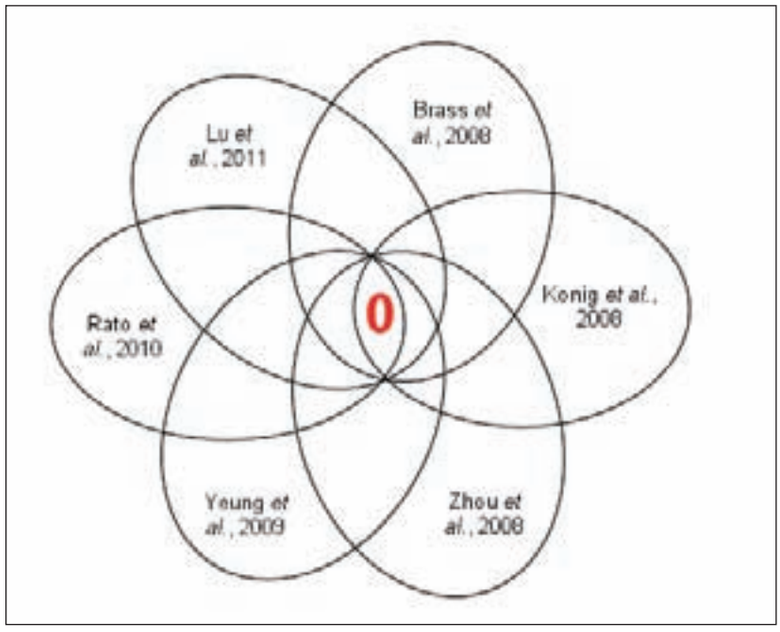

Figure 3
Overlap analysis of six independent RNAi screens performed in order to identify host factors involved in HIV infection with zero commonality

same RNAi library but in a pooled format and performed against 102 cell lines including NOMO-1 cells; once again TBK1 was completely absent and so was STK33[29]. More recently, an academic group from Vanderbilt University (USA) on a pursuit to independently validate TBK1 seems to refute the claims made by Barbie as to the essentiality of TBK1 in KRAS-dependent cancers[35]. Another noticeable example would be PLK1, identified as a target for cancer therapy from RNAi screening efforts undertaken by five different groups and in different cancers[27,36-39]. Such extensive tumour type coverage is suggestive of a more general association with cancer vulnerability. Its identification by RNAi screening methods does not make it a novel target, since it has been studied for more than a decade[40-42]. Besides, PLK1 is also a routinely used positive control in various RNAi screens, alluding to perhaps its more generic role in cellular viability[43-45].

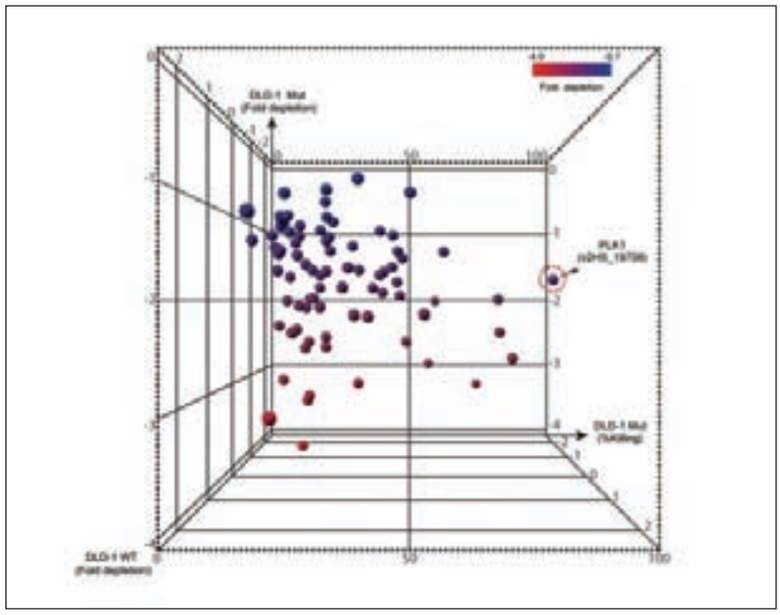

Figure 4: Lack of correlation between fold depletion (pooled shRNA screens) vs cellular viability assessments. 3D correlation plot of the fold depletions in DLD-1 Mut and DLD-1 WT versus percentage cell kill conferred by corresponding hairpins; hairpins with higher depletions in pooled screen failed to reproduce lethal effect on cells in the secondary screen[27]

Remarkably, though very few, we did find some success where other independent research groups have pursued some of the gene targets (**Table 1**). As an example, in 2006 Ngo and co-workers identified CARD11 from a loss of function RNAi screen in diffuse large B-cell lymphoma (DLBCL)[46]. Dong and co-workers followed up this target by using somatic mutagenesis techniques and confirmed that CARD11 was indeed an NF-κB activator in DLBCL[47]. Brass and co-workers identified IFITM proteins as potential targets in host cell infection inflected by influenza A virus H1N1, West Nile Virus, and Dengue Virus from a pooled siRNA screen[26]. Subsequently, Bailey and co-workers were able to validate the role of one of the IFITM proteins, IFITM3, which was shown to reduce H1N1 infection in knockout mouse models[48]. But the fact remains that there is a huge discrepancy in the volume of gene targets allegedly obtained by RNAi screens and the minimal number of targets progressing to independent validation. There must be a reason as to why this has been the outcome, and most likely the culprit may lie in the data analysis leading to hit nomination.

RNAi screening, data analysis and fictitious hits

In 2008, Brass reported on a genome-wide siRNA pooled screen in search of targets to combat HIV infection with a list of 281 genes[21]. Konig published another pooled siRNA screen for factors involved in HIV infection and reported a list of 273 genes; with only 13 overlapping genes with those reported by Brass[22]. Zhou reported performing a screen similar to Brass using similar cell line, screening technology and assay readout; and resulting in a list of 390 genes[23]; a minimal overlap of 15 genes were in common with Brass and 10 genes in common with Zhou. A dismal overlap of only three genes, MED6, MED7 and RELA, were common to the three pooled siRNA screens. In another attempt to identify HIV host factors a year

later, Yeung and co-workers employed a genome-wide shRNA pooled screen and reported a list of 252 genes[49]. In 2010, Rato and co-workers also conducted a pooled shRNA hairpin screen to study the same mechanism using a smaller set of 802 genes measuring reduction in viral expression and identified 14 novel genes, and with zero overlap with the four prior publications[50]. We also evaluated the data outputs from an additional shRNA hairpin screen conducted using a focused library constituted of 94 interferon stimulating genes (ISGs)[51]. Overall, six different RNAi screens to identify host factors for the same HIV virus resulted in zero commonality; such is the current state of RNAi screening – deplorable by any standards that the CCR5 receptor required for HIV infection of host cells could not have been identified (**Figure 3**). Is this hit discordance a common problem for random RNAi screening? To further shed some light as to the extent of this issue, we compared RNAi data outputs from 64 published gene lists from 30 publications reporting on essential genes for cellular viability[52]. Unfortunately, we came to the same conclusion as for the HIV screens, zero gene commonality across the board even for those essential ones[52]. We had postulated that a strong essential gene candidate identified in one screen should reproduce a consistent strong phenotype in others; irrespective of the RNAi technology used. Therefore the observations pertaining to zero overlaps at gene-level remain perplexing[52].

PLK1 was reported as a hit among others by Luo and co-workers in their search for genes which would preferentially kill cells harbouring mutated KRAS[27]. Its nomination was based solely on one active hairpin ranking 118 out of the 379 selected actives. It was also found to be active in their secondary cell kill assay where its overall performance outplayed all other hairpins, an unusual result considering its ranking in the primary pooled screen and subsequent cell kill; and questions the merits of fold depletion as a readout in pooled shRNA screening versus cellular viability as a measure of gene essentiality – clearly no correlation between the two was observed as depicted in **Figure 4**. The activity of the same hairpin was missing from the cell kill assay against the HCT116 cell line harbouring KRAS mutation, though it was reported as active by the same group one year earlier[53]; a perplexing observation of hit discordance within the same research group[27,53]. Furthermore, a close examination of the raw fold change values for the active hairpins revealed an unexpected finding, where 31 shRNA hairpins yielded exactly the same numerical fold change across two independent

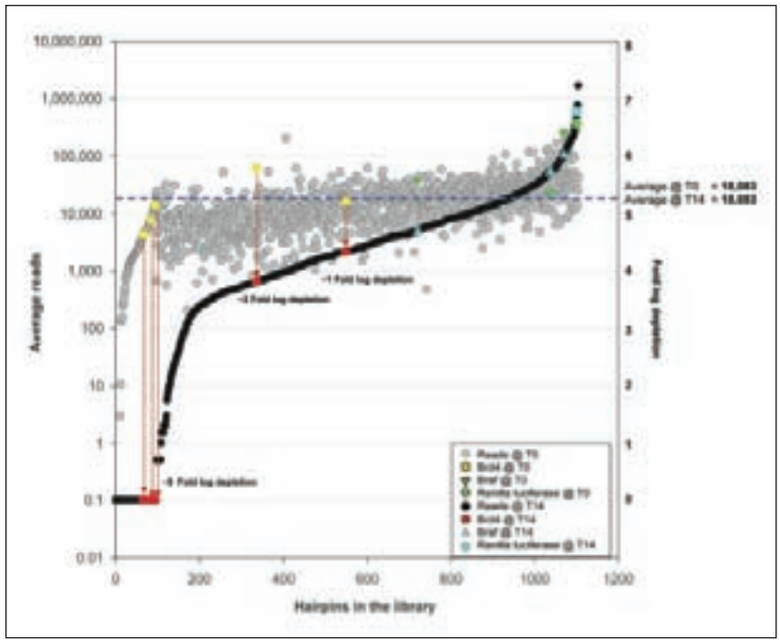

Figure 5: Visualisation of the shRNA pooled screening data leading to the nomination of Brd4 as the only hit. Up to six logs spread in reads at T0 inconsistent with the behaviour of a highly tunable expression vector. Only three hairpins produced 100% depletion @ T14 given that their reads were below average at T0. Hairpin #1448 produced a ~7-fold change @ T14, inconsistent with the 20-fold change determined by Zuber[57]

pooled shRNA screens performed by the same group and at least one year apart (data not shown). Due to the inherent randomness of the process, the multiple integration of hairpins per cell and the inherent noise of the pooled screening approaches, it would be highly unlikely to obtain the same exact fold change activity, even if the screens were to be performed in parallel. Additionally, two genes, DHX8 and LOC149654, were reported as cell kill assay hits in Luo report[27] but shRNA hairpins targeting them were not found among the identified actives in the screen to begin with; yet another perplexing observation this time, fictitious gene hits as a result of random RNAi screening.

The few examples described here concur with the misuse and abuse of the technology and begin to explain, in part, the lack of any breakthroughs with novel gene targets to fight disease – very few needles found thus far; it seems that desperation to publish calls for lower hit nomination standards. Moreover, close scrutiny of multiple RNAi screens reveals an important facet of the results, where the majority of the hits are associated by a maximum number of one active RNAi duplex and this could potentially be the reason for discrepancy of hits at inter-screen level. For example, 56% of the genes reported by Brass meet this criterion[21], 93% in the case of the Luo report[27], and 81% in the case of the Schlabach report[53]. It is also a fact that RNAi screening is abundant in false positives mediated by OTEs, which in some instances have been shown to populate the top ranking hits[54]. The restitute lies in the fact that RNAi is combinatorial in nature and quantification of maximal active duplexes targeting individual genes would serve as a robust measure to assess hit legitimacy. Recently, we reported on a simple comprehensive method and introduced the H score as a hit rate per gene and defined as the total number of active duplexes divided by the total number of duplexes per gene multiplied by 100[20]. This would help address the basic question of the quantitative bias towards cumulative activity relative to inactivity with regards to the targeting duplexes of a given gene, irrespective of the type of screening technology and nature of data output used. In the wake of the current outcry in terms of standardisation of RNAi data analysis practices, a straightforward statistic like the H score would be helpful in scoring for high confidence outcomes in random RNAi screens[20,55-56].

Pooled ShRNA screening: a bridge to nowhere

In a recent comparative analysis of 30 RNAi lethality screens, we unexpectedly observed an unprecedented enrichment in hit genes obtained exclusively from pooled shRNA screening formats; 58% of the total hits were exclusively from pooled shRNA hairpin screens[52]. Clearly shRNA pooled screens, viewed by many researchers as cheap and quick, have become a reservoir of targets waiting judgment day. On the other hand, in some instances, the magical 'funnel' singles out one high value target from a shRNA hairpin screen, which is emphasised to hold therapeutic potential in curing disease. As an example, Brd4 was recently reported as the only strong hit from a random pooled shRNA hairpin screen[57]; in total contradiction to many of these types of screens where hundreds of hits were found and reported – describing perhaps the overall inherent issue of the pooled approaches. It is also rather perplexing that one gene target gets identified amidst a pooled shRNA screen which has a very high dynamic range and, therefore, a huge noise associated with it as depicted in **Figure 5**. Such variation in hits, quantitative as well as claimed qualitative, indeed emphasises a serious note of caution during hit nomination, more so in lethality screens raising suspicions as to whether a lethal phenotype is indeed indicative of a target's essentiality in cell

survival or is merely an outcome of random noise. However, if there were no small molecule inhibitors for Brd4 available, would this group have singled out Brd4 as the sole hit from their pooled shRNA screen? The most likely answer would have been no.

Despite technological advances in shRNA screening over the past decade[58-60], it is a matter of serious concern that the issues of random noise and heterogeneity remain unaddressed. Perhaps it is because of the four key determinants of shRNA hairpin functionality which remain a black box: 1) Efficient delivery of DNA-based plasmid into the host cell; 2) Followed by its random and hopefully stable genomic integration; 3) Its successful transcription to yield pri-shRNA hairpin; and 4) Precise processing of pre-shRNA hairpin to yield a functional silencing duplex; in addition to producing a measurable degree of phenotypic perturbation as a result of knocking down the intended target. Gu and co-workers have recently provided the first experimental evidence in support of uncertainty associated with hairpin processing inside the cell; multiple oligonucleotide strands of varying lengths were obtained as cleavage products[61]. This evidence along with previous concerns over DICER-mediated specificity of hairpin cleavage is indicative of the inefficiencies in intracellular hairpin cleavage and subsequent origin of incognito OTEs in shRNA hairpin screens[61-63].

Taking a step further, it also becomes critical to evaluate whether this differential hairpin processing is generic, cell line or technology specific. In a recent genome-wide shRNA hairpin lethality screen for genes essential in cellular viability[28], PLK1 was expected to be an obvious strong hit. The screened TRC1 library contained 23 hairpins targeting PLK1, out of which 20 hairpins were deemed independently validated[28], but only four hairpins scored as active and causing cell death. Similar trends with regards to validated hairpins were also observed in the case of several other known regulators of cell survival, like AURKA, and WEE1, to name a few[28]. Importantly, the cell lines used in screening and in the validation process of the hairpins are more often distinct. In another example provided by Boettcher and co-workers, variable hairpin activity in terms of knockdown efficiencies was observed in two different cell lines, albeit hairpins used here were designed based on the miR-30 backbone-derived shRNA technology[64]. Such findings are suggestive of a correlation between phenotypic perturbations produced by individual shRNA hairpins and the cell lines used; perhaps differential hairpin processing is indeed cell line specific, and we should take it into account when comparing shRNA results using different cell lines.

There are two additional aspects exclusive to the pooled shRNA screens, those of synergy and data deconvolution. A pool could be comprised of up to hundreds of thousands of hairpins, targeting a variety of genes collectively as opposed to a relatively controlled one-gene one-hairpin environment as is the case for arrayed formats. Variable transduction efficiencies lead to heterogeneous hairpin representation(s) per cell[65], and the lack of definitive experimental checkpoints to control for these transactions leave considerable room to speculate zero to many hairpin integrations per individual cell. Also taking into consideration the intracellular equilibrium orchestrated by the endogenous RNAi (**Figure 1**), how can we ascertain that the observed phenotypic perturbation is attributed to a single target knockdown and not a synergy concerted by multiple distinct hairpins? As for data deconvolution, PCR amplification of the extracted genomic DNA is the primary step of the workflow and PCR amplification are prone to errors[66], which serve as the first port of entry for noise post-screening. Moreover, the two commonly used technologies of hairpin retrieval, that of deep sequencing and microarray hybridisation, fall prey to sequencing errors, a second port of entry for the noise in the data output[65]; clearly additional experimental noise is introduced during data deconvolution of pooled shRNA hairpin screens. Therefore, it would be reasonable to question the merits of performing pooled shRNA screening, and the value proposition associated with the resulting hits, especially when their outcome is null (**Table 1**).

Core concerns and considerations

In recent years, serious and valid concerns about the gene targets identified through this random RNAi screening process have begun to emerge; thus it is not surprising that no outcome is in sight (**Table 1**) as one would have predicted based on the sheer enormity of the combined work leading to ~580 published RNAi screens. The first signs of trouble for the technology came about in 2008, when four independent research groups performed RNAi screens in order to identify novel host factors modulating HIV entry and replication; they published completely different gene target lists with zero overlap; not even the critical host receptor, CCR5, upon which the virus relies for successful entry into the host cell (see above). Protectors of this noble technology rushed in very quickly claiming difficulties in comparing results across these four screens[25], but missed a golden

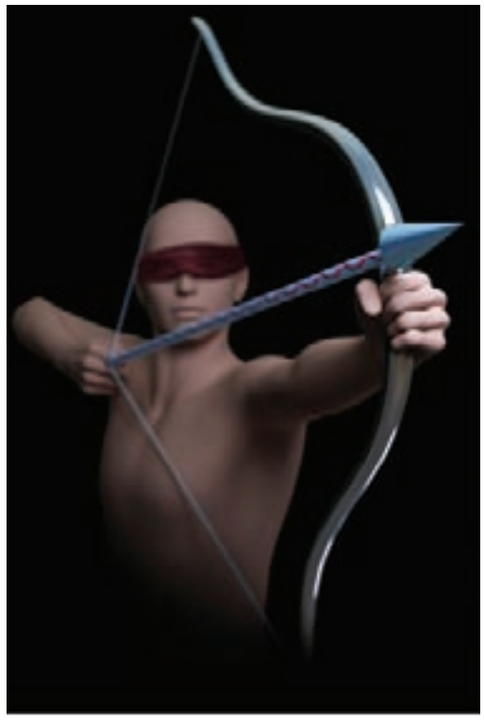

Figure 6
Random RNAi vis-à-vis blindfolded archery; games of luck and a magnificent quest for elusive targets. A blindfolded archer hurls an arrow to hit bull's-eye, likewise a scientist seeks RNAi in pursuit of a target; ushered by unquestioned faith, both stand irresolute as to the final outcomes of their craft. *(Illustration by Wenjing Wu)*

opportunity to use this disappointing outcome to address the technology's major caveats and pitfalls; and more so, set the foundations for a stringent and robust list of standards and guidelines for: 1) assay development strategies and feasibility, 2) RNAi cargo delivery and efficiency versus viability, and finally, 3) data analysis and gene hit nomination. Data reproducibility is the very lifeline of scientific research, and the field of RNAi is no exception. Top hits resulting from pitfalls of passion that failed to withstand the test of reproducibility in cell-based assays leaves room for remedial effort, inexact identification in essential gene status in patient-based shRNA screening studies would bring to bear far more profound ramifications: It may pose a direct impact on prognostics and patient well-being. We have already witnessed that RNAi is not immune to hit discordance, retraction of published work, and unfortunately fraudulent claims pertaining to discoveries made through its use[24,25,32-35,67-71]. As an example, Lipardi and Paterson took the honourable course of action by retracting their RNAi screening paper due to misinterpretation of their screening data analysis[67]. Amidst all these core concerns, we are losing faith in the merits of novel gene candidates identified through this random RNAi screening process and who would step in to validate them.

If this technology is to succeed and flourish, there is a daunting task in front of us to standardise the way we develop, execute and interpret RNAi screening data output; perhaps data transparency, open data sharing especially in terms of raw values, cross-screen comparisons based on these raw data outputs from different research groups, and the use of simple statistics like the H score to better nominate hits, would be an ideal starting point; in turn empowering the peer-review process to require stringent criteria for hit nomination to shape up the outcome of published gene lists; by rejecting any gene hit nominated based on one active out of three siRNA duplexes or one to two actives out of five shRNA hairpins – a hit must have an H score >60. This stringency would potentially help restore the much needed faith in this technology to withstand the test of time, in search of the elusive target(s) (**Figure 6**), and would set the stage for standardising the meaning of a hit nominated from random RNAi screening.

Concluding remarks

As the dust settles on these thousands of genes identified using random RNAi screening, one but wonders how many are indeed 'legitimate' to begin with, the data of which can easily be reproduced by others, and would prevail with a positive therapeutic outcome to fight global disease at a time where we are in very desperate need of novel targets for small molecule intervention. As noble a technology as RNAi may well be, it does not mean it is above scientific scrutiny, nor immune to artifacts and abuse, and some may even hold the view that it is simply a means to an end. We are cautiously optimistic that we will overcome its flaws through teaching, data transparency and stringent standards for hit nomination; as the novelty of its bells and whistles has begun to wear off.

Acknowledgements

We wish to thank members of the HTS lab and colleagues worldwide for fruitful discussions on the merits of RNAi screening technologies. The HTS Lab is partially supported by William H. Goodwin and Alice Goodwin and the Commonwealth Foundation for Cancer Research, the Experimental Therapeutics Center of the Memorial Sloan-Kettering Cancer Center, the William Randolph Hearst Fund in Experimental Therapeutics, the Lillian S. Wells Foundation and by a NIH/NCI Cancer Center Support Grant 5 P30 CA008748-44. DDW

***Dr Hakim Djaballah**, molecular pharmacologist and technologist, has been the Director of the HTS*

Core Facility at Memorial Sloan-Kettering Cancer Center since its establishment in 2003. In 1992, he received his PhD in biochemistry from the University of Leicester, England. He was the recipient of the 2007 Robots and Vision User Recognition Award.

***Bhavneet Bhinder** is a computational analyst at the HTS Core Facility who was one of the pioneers of the BDA methodology to analyse RNAi screening data across all technology platforms and formats. She joined the group in 2009. In 2008, she received her MS in Bioinformatics from Georgia Institute of Technology, USA.*

References

1 Guo, S and Kemphues, KJ. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81, 611-20 (1995).
2 Mohr, S et al. Genomic screening with RNAi: results and challenges. Annu Rev Biochem 79, 37-64 (2010).
3 Jackson, AL and Linsley, PS. Recognizing and avoiding siRNA OTEs for target identification and therapeutic application. Nat Rev Drug Discov 9, 57-67 (2010).
4 Sigoillot, FD and King, RW. Vigilance and validation: Keys to success in RNAi screening. ACS Chem Biol 6, 47-60 (2011).
5 Mattick, JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO reports 2, 986-91 (2001).
6 Burge, SW et al. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41, D226-32 (2013).
7 Friedman, RC et al. Most mammalian mRNAs are conserved targets of MicroRNAs. Genome Res 19, 92-105 (2009).
8 Batista, PJ and Chang, HY. Long noncoding RNAs: cellular address codes in development and disease. Cell 152, 1298-307 (2013).
9 International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. Nature 431, 931-45 (2004).
10 Pennisi, E. Gene Counters Struggle to Get the Right Answer. Science 301, 1040-1 (2003).
11 Lord, CJ et al. RNA interference screening demystified. J Clin Pathol 62, 195-200 (2009).
12 Falschlehner, C et al. High-throughput RNAi screening to dissect cellular pathways: a how-to guide. Biotechnol J 5, 368-76 (2010).
13 Hong-Geller, E and Micheva-Viteva, SN. Functional gene discovery using RNA interference-based genomic screens to combat pathogen infection. Curr Drug Discov Technol 7, 86-94 (2010).
14 Quon, K and Kassner, PD. RNA interference screening for the discovery of oncology targets. Expert Opin Ther Targets 13, 1027-35 (2009).
15 Mizuarai, S and Kotani, H. Synthetic lethal interactions for the development of cancer therapeutics: biological and methodological advancements. Hum Genet 128, 567-75 (2010).
16 Swanton, C et al. Predictive biomarker discovery through the parallel integration of clinical trial and functional genomics datasets. Genome Med 2, 53 (2010).
17 Wang, XZ and Weaver, DT. The ups and downs of DNA repair biomarkers for PARP inhibitor therapies. Am J Cancer Res 1, 301-27 (2011).
18 Alberts, B et al. Studying Gene Expression and Function. Molecular Biology of the Cell 4th edition. Garland Science, New York (2002).
19 Goldsmith, ZG and Dhanasekaran, N. The microrevolution: applications and impacts of microarray technology on molecular biology and medicine (review). Int J Mol Med 13, 483-95 (2004).
20 Bhinder, B and Djaballah, H. A Simple Method for Analyzing Actives in Random RNAi Screens: Introducing the "H Score" for Gene Nomination & Prioritization. Comb Chem High Throughput Screen 15, 686-704 (2012).
21 Brass, AL et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science 319, 921-6 (2008).
22 König, R et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135, 49-60 (2008).
23 Zhou, H et al. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe 4, 495-504 (2008).
24 Bushman, FD et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS Pathog 5, e1000437 (2009).
25 Pache, L et al. Identifying HIV-1 host cell factors by genome-scale RNAi screening. Methods 53, 3-12 (2011).
26 Brass, AL et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell 139, 1243-54 (2009).
27 Luo, JL et al. Genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 137, 835-48 (2009).
28 Bhinder, B et al. An arrayed genome-scale lentiviral-enabled short hairpin RNA screen identifies lethal and rescuer gene candidates. Assay Drug Dev Technol 11,173-90 (2013).
29 Cheung, HW et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci USA 108, 12372-77 (2011).
30 Scholl, C et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. Cell 137, 821–34 (2009).
31 Barbie, DA et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 462, 108-12 (2009).
32 Naik, G. Scientists' Elusive Goal: Reproducing Study Results. WSJ, online.wsj.com (2011).
33 Babij, C et al. STK33 kinase activity is non-essential in KRAS-dependent cancer cells. Cancer Res 71, 5818-26 (2011).
34 Luo, T et al. STK33 kinase inhibitor BRD-8899 has no effect on KRAS-dependent cancer cell viability. Proc Natl Acad Sci U S A 109, 2860-5 (2012).
35 Vangamudi, B et al. Evaluation of TBK1 as a novel cancer target in the K-Ras pathway. Cancer Res 72, Abstract nr 1813 (2012).
36 Watt, SA et al. Integrative mRNA profiling comparing cultured primary cells with clinical samples reveals PLK1 and C20orf20 as therapeutic targets in cutaneous squamous cell carcinoma. Oncogene 30, 4666-77 (2011).
37 Ding, Y et al. Combined gene expression profiling and RNAi screening in clear cell renal cell carcinoma identify

PLK1 and other therapeutic kinase targets. Cancer Res 71, 5225-34 (2011).
38 Duan, Z et al. Lentiviral shRNA screen of human kinases identifies PLK1 as a potential therapeutic target for osteosarcoma. Cancer Lett 293, 220-9 (2010).
39 Hu, K et al. Small interfering RNA library screen identified polo-like kinase-1 (PLK1) as a potential therapeutic target for breast cancer that uniquely eliminates tumor-initiating cells. Breast Cancer Res 14, R22 (2012).
40 Wolf, G et al. Polo-like kinase: a novel marker of proliferation: correlation with estrogen-receptor expression in human breast cancer. Pathol Res Pract 196, 753-9 (2000).
41 Takahashi, T et al. Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. Cancer Sci 94, 148-52 (2003).
42 Tokumitsu, Y et al. Prognostic significance of polo-like kinase expression in esophageal carcinoma. Int J Oncol 15, 687-92 (1999).
43 Sarthy, AV et al. Survivin depletion preferentially reduces the survival of activated K0-Ras-transformed cells. Mol Cancer Ther 6, 269-76 (2007).
44 Zheng, M et al. Growth inhibition and radiosensitization of glioblastoma and lung cancer cells by small interfering RNA silencing of tumor necrosis factor receptor-associated factor 2. Cancer Res 68, 7570-8 (2008).
45 Cole, KA et al. RNAi screen of the protein kinome identifies checkpoint kinase 1 (CHK1) as a therapeutic target in neuroblastoma. Proc Natl Acad Sci USA 108, 3336-41 (2011).
46 Ngo, VN et al. A loss-of-function RNA interference screen for molecular targets in cancer. Nature 441, 106-10 (2006).
47 Dong, G et al. A20, ABIN-1/2, and CARD11 mutations and their prognostic value in gastrointestinal diffuse large B-cell lymphoma. Clin Cancer Res 17, 1440-51 (2011).
48 Bailey, CC et al. Ifitm3 limits the severity of acute influenza in mice. PLoS Pathog 8, e1002909 (2012).
49 Yeung, ML et al. A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. J Biol Chem 284, 19463-73 (2009).
50 Rato, S et al. Novel HIV-1 knockdown targets identified by an enriched kinases/phosphatases shRNA library using a long-term iterative screen in Jurkat T-cells. PLoS One 5, e9276 (2010).
51 Lu, J et al. The IFITM proteins inhibit HIV-1 infection. J Virol 85, 2126-37 (2011).
52 Bhinder, B and Djaballah, H. Systematic analysis of RNAi reports finds zero gene commonality & reveal an unprecedented enrichment in pooled screens: does depletion warrant essentiality? Assay Drug Dev Technol (2013) (under review).
53 Schlabach, MR et al. Cancer proliferation gene discovery through functional genomics. Science 319, 620-4 (2008).
54 Schultz, N et al. Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-β pathway and reveal microRNA regulation of TGFBR2. Silence 2, 3 (2011).
55 Djaballah, H. Editorial: Random RNAi Screening Data Analysis: A Call for Standardization. Comb. Chem. High Throughput Screen 15, 685 (2012).
56 Kaelin Jr, WG. Use and Abuse of RNAi to study Mammalian Gene Function. Science 337, 421-2 (2012).
57 Zuber, J et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature 478, 524-8 (2011).
58 Berns, K et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 428, 431-7 (2004).
59 Moffat, J et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124, 1283-98 (2006).
60 Chang, K et al. Lessons from nature: microRNA-based shRNA libraries. Nat Methods 3, 707-14 (2006).
61 Gu, S et al. The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. Cell 151, 900-11(2012).
62 Park, JE et al. Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature 475, 201-5 (2011).
63 Vermeulen, A et al. The contributions of dsRNA structure to Dicer specificity and efficiency. RNA 11, 674-82 (2005).
64 Boettcher, M and Hoheisel, JD. Pooled RNAi Screens – Technical and Biological Aspects. Curr Genomics 11, 162-7 (2010).
65 Sims, D et al. High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. Genome Biol 12, R104 (2011).
66 Kobayashi, N, Tamura, K and Aotsuka T. PCR error and molecular population genetics. Biochem Genet 37, 317-21 (1999).
67 Lipardi, C and Paterson, BM. Retraction for Lipardi and Paterson. Identification of an RNA-dependent RNA polymerase in Drosophila involved in RNAi and transposon suppression. Proc Natl Acad Sci U S A 108, 15010 (2011).
68 Zhang, S et al. Retracted manuscript: Overexpression of transketolase protein TKTL1 is associated with occurrence and progression in nasopharyngeal carcinoma. Cancer Biol Ther 10, 110 (2010).
69 Department Of Health And Human Services: Findings of Scientific Misconduct. www.gpo.gov/fdsys/pkg/FR-2010-08-25/html/2010-21048.htm (accessed May 16, 2013).
70 Dolgin, E. Iowa biologist falsified figures. The Scientist, www.the-scientist.com (2008).
71 Ledford, H. Drug giants turn their backs on RNA interference. Nature 468, 487 (2010).
72 Banerji, V et al. The intersection of genetic and chemical genomic screens identifies GSK-3α as a target in human acute myeloid leukemia. J Clin Invest 122, 935-47 (2012).
73 Ebert, BL et al. Identification of RPS14 as a 5q-syndrome gene by RNA interference screen. Nature 451, 335-9 (2008).
74 Czibere, A et al. Low RPS14 expression is common in myelodysplastic syndromes without 5q- aberration and defines a subgroup of patients with prolonged survival. Haematologica 94, 1453-5 (2009).
75 Herrmann, H et al. Small-molecule inhibition of BRD4 as a new potent approach to eliminate leukemic stem- and progenitor cells in acute myeloid leukemia AML. Oncotarget 3, 1588-99 (2012).
76 Gobeil, S et al. A genome-wide shRNA screen identifies GAS1 as a novel melanoma metastasis suppressor gene. Genes Dev 22, 2932-40 (2008).

77 Wajapeyee, N et al. Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. Cell 132, 363-74 (2008).
78 Scurr, LL et al. IGFBP7 is not required for B-RAF-induced melanocyte senescence. Cell 141, 717-27 (2010).
79 Schrama, D et al. Lack of correlation between IGFBP7 expression and BRAF mutational status in melanoma. J Invest Dermatol 130, 897-8 (2010).
80 Boehm, JS et al. Integrative genomic approaches identify IKBKE as a breast cancer oncogene. Cell 129, 1065-79 (2007).
81 Hölzel, M et al. NF1 is a tumor suppressor in neuroblastoma that determines retinoic acid response and disease outcome. Cell 142, 218-29 (2010).
82 Kolfschoten, IG et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell 121, 849-58 (2005).
83 Westbrook, TF et al. A genetic screen for candidate tumor suppressors identifies REST. Cell 121, 837-48 (2005).
84 Eichhorn, PJ et al. USP15 stabilizes TGF-β receptor I and promotes oncogenesis through the activation of TGF-β signaling in glioblastoma. Nat Med 18, 429-35 (2012).
85 Porter, CC et al. Integrated genomic analyses identify WEE1 as a critical mediator of cell fate and a novel therapeutic target in acute myeloid leukemia. Leukemia 26, 1266-76 (2012).
86 Vicent, S et al. Wilms tumor 1 (WT1) regulates KRAS-driven oncogenesis and senescence in mouse and human models. J Clin Invest 120, 3940-52 (2010).
87 Geiger, R et al. BAP31 and BiP are essential for dislocation of SV40 from the endoplasmic reticulum to the cytosol. Nat Cell Biol 13, 1305-14 (2011).
88 Krishnan, MN et al. RNA interference screen for human genes associated with West Nile virus infection. Nature 455, 242-5 (2008).
89 Fernandez-Garcia, MD et al. Appraising the roles of CBLL1 and the ubiquitin/proteasome system for flavivirus entry and replication. J Virol 85, 2980-9 (2011).
90 Swanton, C et al. Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs. Cancer Cell 11, 498-512 (2007).
91 Lupberger, J et al. EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. Nat Med 17, 589-95 (2011).
92 Giamas, G et al. Kinome screening for regulators of the estrogen receptor identifies LMTK3 as a new therapeutic target in breast cancer. Nat Med 17, 715-9 (2011).
93 Stebbing, J et al. LMTK3 expression in breast cancer: association with tumor phenotype and clinical outcome. Breast Cancer Res Treat 132, 537-44 (2012).
94 Bajrami, I et al. Synthetic lethality of PARP and NAMPT inhibition in triple-negative breast cancer cells. EMBO Mol Med 4, 1087-96 (2012).
95 Santidrian, AF et al. Mitochondrial complex I activity and NAD+/NADH balance regulate breast cancer progression. J Clin Invest 123, 1068-81 (2013).
96 Morgan-Lappe, SE et al. Identification of Ras-related nuclear protein, targeting protein for xenopus kinesin-like protein 2, and stearoyl-CoA desaturase 1 as promising cancer targets from an RNAi-based screen. Cancer Res 67, 4390-8 (2007).
97 Zhang, H et al. Scd1 Plays a Tumor-Suppressive Role in Survival of Leukemia Stem Cells and the Development of Chronic Myeloid Leukemia. Mol Cell Biol 32, 1776-87 (2012).
98 Aguirre-Portolés, C et al. Tpx2 controls spindle integrity, genome stability, and tumor development. Cancer Res 72, 1518-28 (2012).
99 Szász, AM et al. The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. PLoS One 8, e56707 (2013).