Phenovault: an open-access resource for analysing RNAi and CRISPR screens

Catherine Goh, Andrew Walsh, Michaela Beitzinger, Stefan Hannus, Michael Hannus
siTOOLs Biotech GmbH, Lochhamer Strasse 29A, 82152 Martinsried/Planegg, Germany

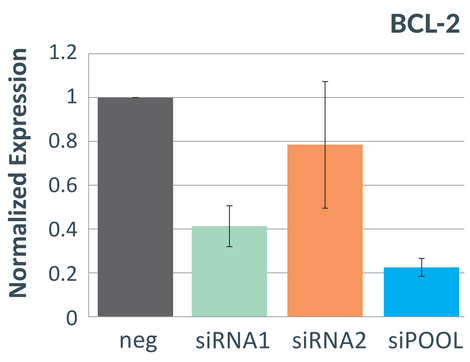
RNAi and CRISPR screens give rise to a wealth of information on gene function. Frequently, however, published analysis focuses on a few top-scoring genes. Consequently, complete genome-scale screening data often lies buried within supplementary materials, invisible and untapped. The Phenovault is a growing database and analysis suite hosted by siTOOLs Biotech that contains complete, reagent sequence/ID-linked datasets from published RNAi/CRISPR screens. With over 20 million data-points, the Phenovault is the largest curated RNAi screening repository. Together with public and proprietary algorithms that harness the dominant microRNA seed-based behaviour of siRNAs, the Phenovault helps researchers uncover novel insights on their 1) siRNA reagents, 2) target genes and 3) RNAi screening datasets. For updates, visit: www.phenovault.de.

siRNA **reagent** evaluation

Predict siRNA off-target activity and identify off-target genes

Validation of off-target activity for BCL2 siRNA/siPOOL:
1. A549 cells were transfected with two different single siRNAs or siPOOL against BCL2 at 3 nM.
2. TargetScan (Lewis et al., 2015) was performed to identify candidate off-target genes.
3. Real-time qPCR measurement of on-target and off-target RNA levels was performed on the same RNA sample, collected 24 h post-treatment.

On-target activity



Single siRNAs showed variable on-target silencing activity (0-60%) and 80% knock-down (KD) seen with siPOOL.

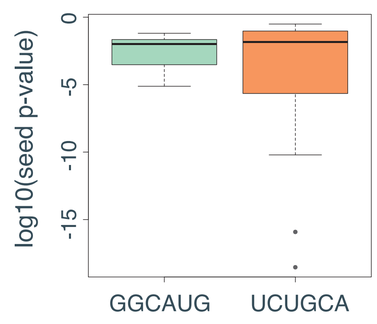
Off-target activity

Reagent	Candidate off-targets tested	Candidate off-targets > 50% KD	Sig. of findings, p-value*
siRNA1	12	1	0.11
siRNA2	5	2	0.00096
siPOOL	7	0	0.93

**P-values calculated on assumption that 100 out of 10,000 expressed genes would show > 50% KD by random chance.*

siRNA2 showed significant off-target activity as compared to siRNA1 and siPOOL.

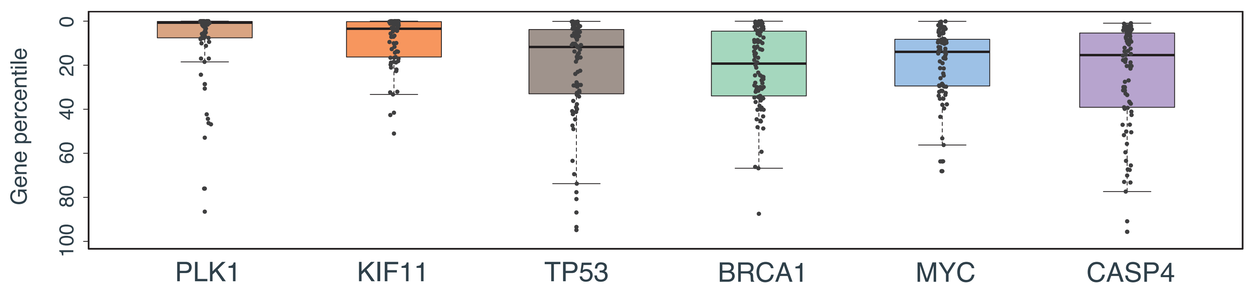
Predicted off-target activity (Phenovault)



Greater seed enrichment of siRNA2 in top hits of Phenovault screens indicate high off-target activity.

Target **gene** evaluation

Reveal functional phenotypic features associated with gene targets



Each point represents a screening feature. Gene percentiles were calculated considering up to three screening features (readouts) from each Phenovault screen. For pooled-siRNA screens, actual feature values were used to calculate gene percentiles. For single-siRNA screens, gene percentiles were calculated using the RSA p-value for the gene. This was calculated using all siRNAs for the gene with the RSA algorithm (König et al., 2007).

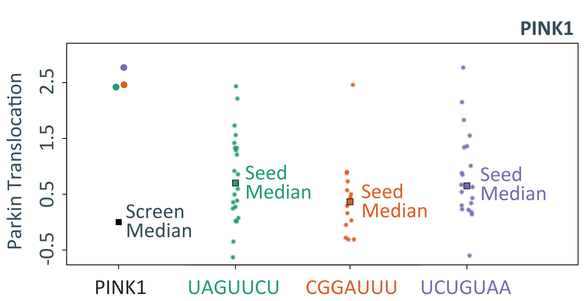
A gene percentile of 20 means that the gene is ranked in the top 20% of genes for that feature.
Screening features relate to phenotypes such as cell viability, cell cycle arrest, endocytosis and more.
Phenovault retrieves a list of relevant phenotypic features associated with target gene(s) of interest and publications linked to those screens.

RNAi screen **dataset** re-analysis with seed-based focus

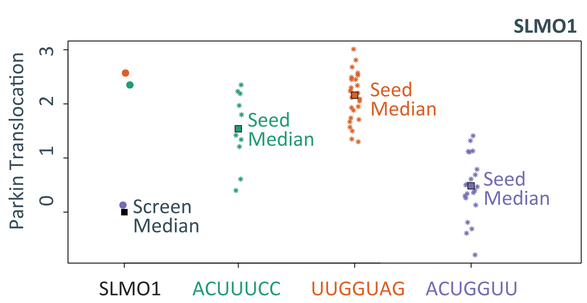
Identify false positives and uncover novel hits

An RNAi screen (3 siRNAs tested per gene) for factors regulating Parkin mitochondrial translocation was analysed (Hasson et al., 2013). In HeLa cells, 83,000 siRNAs were screened against 21,993 unique genes at 20 nM for 48h.

PINK1 and SLMO1 were identified as top hits. A seed triage analysis examines the performance of siRNAs sharing the same seed sequences as hit gene siRNAs in the screen. The hit is likely a false positive if siRNAs with same seeds are giving significant read-outs, as is the case for SLMO1.

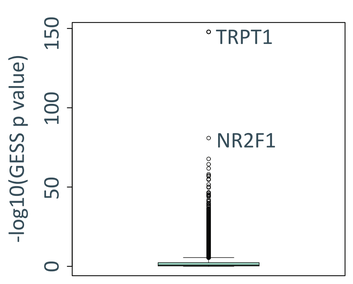


Minimal seed enrichment amongst 3 hit siRNAs indicates PINK1 is likely a true hit.



Significant seed enrichment for 2 of 3 hit siRNAs indicates SLMO1 is likely a false positive.

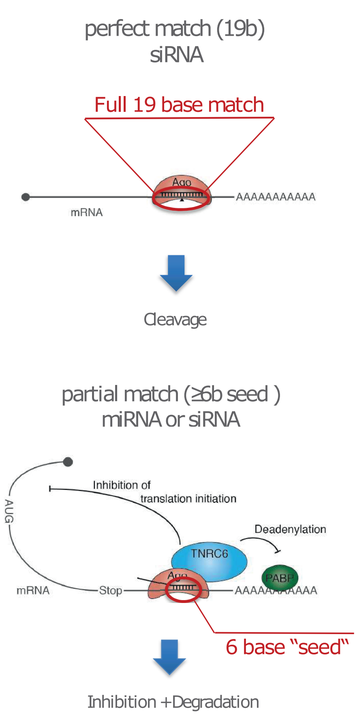
A seed-based hit finding algorithm, GESS identifies novel hits based on seed match-multiplicity to 3'UTR regions (Sigoillot et al., 2012).



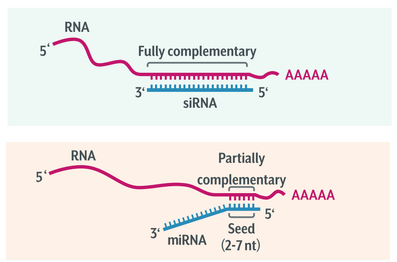
The GESS algorithm identified TRPT1 and NR2F1 as novel hits.

siRNA off-targeting

Mechanism



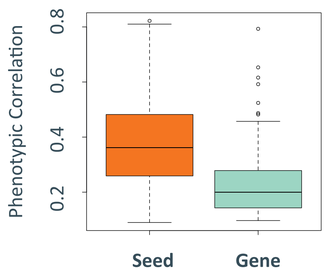
What is a seed?



The seed is a 6 base sequence at position 2 to 7 of the siRNA guide strand that dictates siRNA off-target activity based on microRNA mimicry.

Dominance

Seed effects dominate phenotypes



Correlation analysis of 300 phenotypic features show siRNAs sharing the same seed produce more closely related phenotypes than siRNAs targeting the same gene.

References

Lewis et al. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell. 120, 15–20

Sigoillot, F. D et al. (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. Nat Meth. 9, 363–366

König et al. (2007) A probability-based approach for the analysis of large-scale RNAi screens. Nat. Methods. 4, 847–849

Hasson et al. (2013) High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. Nature. 504, 291–295