Phenovault: RNAi/CRISPR Database & Analysis Suite Uncovers New Insights into Genes & Reagents

siTOOLs BIOTECH

Catherine Goh, Andrew Walsh, Michaela Beitzinger, Jonas Bertram, Michael Hannus

siTOOLs Biotech GmbH, Lochhamer Strasse 29A, 82152 Martinsried/Planegg, Germany
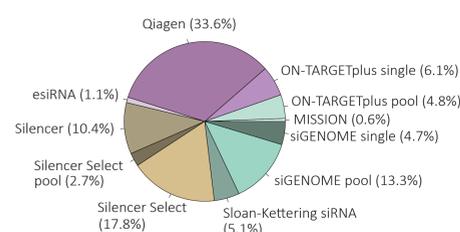
RNAi and CRISPR screens give rise to a wealth of information on gene function. Frequently, however, published analysis focuses on a few top-scoring genes. Consequently, complete genome-scale screening data often lies buried within supplementary materials, invisible and untapped. The Phenovault is a growing 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 uncovel novel insights on their 1) siRNA reagents, 2) target genes and 3) RNAi screening datasets. For updates, visit: www.phenovault.de.
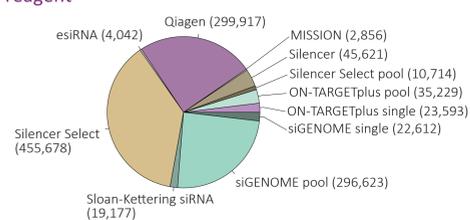
What is Phenovault?

Largest Collection of RNAi Screening Data and RNAi Analysis Algorithms

The Phenovault is a carefully curated database and analysis suite, largely focussed on RNAi screens. By harnessing previously obtained genome-wide functional screening data, coupled with a wide array of published and proprietary data analysis algorithms, the Phenovault serves to maximise our understanding of gene function and gene silencing reagents. As siRNAs are well-established to produce phenotypes via off-target seed-based mechanisms, the Phenovault also uncovers an interesting perspective and novel insights into gene function through siRNA off-target analysis.
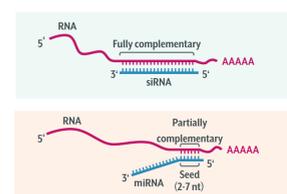

Phenovault RNAi reagents: Qiagen (33.6%), ON-TARGETplus single (6.1%), ON-TARGETplus pool (4.8%), MISSION (0.6%), siGENOME single (4.7%), siGENOME pool (13.3%), Sloan-Kettering siRNA (5.1%), Silencer Select (17.8%), Silencer Select pool (2.7%), Silencer (10.4%), esiRNA (1.1%)

Screen count per Phenovault RNAi reagent: Qiagen (299,917), MISSION (2,856), Silencer (45,621), Silencer Select pool (10,714), ON-TARGETplus pool (35,229), ON-TARGETplus single (23,593), siGENOME single (22,612), siGENOME pool (296,623), Sloan-Kettering siRNA (19,177), Silencer Select (455,678), esiRNA (4,042)
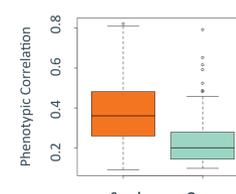
What is a seed?

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.

Seed effects dominate phenotypes

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

Applications

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

The off-target activity of siRNAs can be predicted by examining its seed activity in previous Phenovault screens. Seeds that are „highly active" appear often in siRNAs that are top-ranked across many RNAi screens. The premise being, that top-ranked siRNAs tend to hit multiple genes and produce more significant phenotypes. An siRNA with a „highly active" seed is therefore more likely to produce more off-target effects. We tested this hypothesis with siRNAs targeting BCL-2.

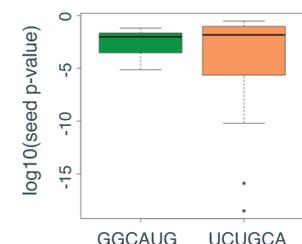**Method:**

2 siRNAs targeting BCL2: siRNA 1, siRNA2

Predict off-target behaviour based on seed activity in Phenovault screens → Predict off-target gene candidates with TargetScan (Lewis et al. 2005)

Test actual off-target behaviour → Analyse expression of off-target genes by rtqPCR after siRNA treatment
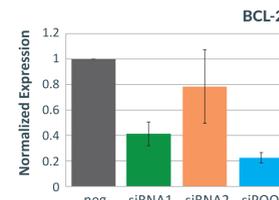
Predicted off-target activity (Phenovault)

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

Actual off-target activity

Single siRNAs showed variable on-target silencing activity (0-60%) and 80% knock-down (KD) seen with siPOOL. siRNA2 showed significant off-target activity as compared to siRNA1 and 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.

Gene Evaluation: Reveal functional phenotypic features associated with gene targets
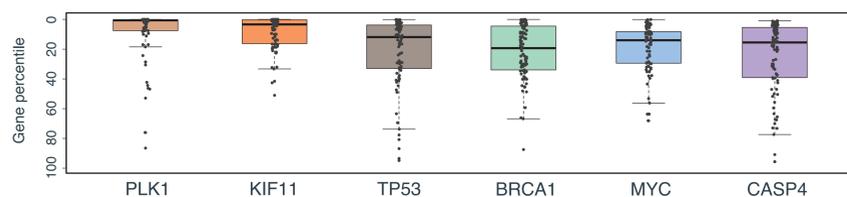

Phenovault's large collection of phenotypes associated with gene knock-downs/outs provide a wealth of information relating to gene function. Phenovault retrieves a list of relevant phenotypic features associated with target gene(s) of interest and publications linked to those screens.

Each point represents a screening feature/read-out. Gene percentiles were calculated based on up to three screening features per Phenovault screen. A gene percentile of 20 means that the gene is ranked in the top 20% of genes for that feature. Gene percentiles calculated based on actual feature values or RSA p-values (König et al., 2007).

RNAi screen dataset re-analysis: Identify false positives and uncover novel hits

Comprehensive on-target and off-target analysis algorithms in Phenovault coupled with siTOOLS Biotech's RNAi screening expertise helps scientists uncover new insights into existing functional genomic screening data. Here, we analysed an RNAi screen performed for factors regulating Parkin mitochondrial translocation (Hasson et al., 2013).

Seed triage analysis identifies false positives in hit list
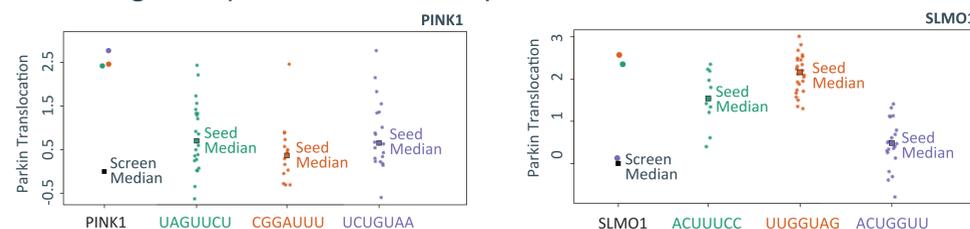
A seed triage analysis examines the phenotypic performance of siRNAs sharing the same seed sequences as top-ranked siRNAs in a screen. If a seed-dependent effect is observed, these top-ranked siRNAs are acting through an off-target based mechanism as opposed to an on-target gene knock-down. This was the case for SLMO1.

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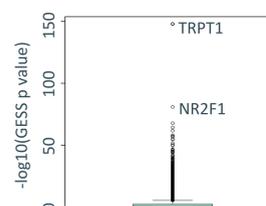
A seed-based hit finding algorithm, GESS identifies novel hits based on seed match-multiplicity to 3'UTR regions (Sigoillot et al., 2012).

The GESS algorithm identified TRPT1 and NR2F1 as novel hits in the Hasson et al. screen.

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