Rapid drug target ranking system for Drug Discovery: A systematic analysis of cancer genomic data from the Oncomine® Knowledgebase identifies an oncogenic role for the NFE2L2 pathway in multiple cancer types

Abstract #2779

Sean Eddy, Mary Ellen Urick, Mark Tomilo, Armand Bankhead, Dan Rhodes, and Emma Bowden, Thermo Fisher Scientific, 110 Miller Avenue, Ann Arbor, MI, USA, 48104

ABSTRACT

Next-generation sequencing (NGS) is poised to accelerate the discovery of new drug targets through identification of genomic markers of response and identification of models to characterize candidate drivers. To maximize the value of NGS, it is imperative to develop data analysis/interpretation solutions that accurately assess genomic aberrations, delineate driver alterations from passengers, annotate alterations for potential clinical relevance and integrate alterations by gene and pathway.

Here, we present our potential framework for the systematic analysis of thousands of NGS samples as well as expertly curated oncology data for the purpose of identifying potential drug targets. Our methodology was developed through a systematic interrogation of genomic aberrations in a training set of gold standard oncogenes such as *EGFR* and *PIK3CA*, and tumor suppressors such as *TP53* and *PTEN*. From a research perspective, the resulting platform was used to rank genes through an assessment of driver genomic aberrations, associations with patient survival, and potential actionability. Using this framework, we found supporting evidence implicating *NFE2L2* as an oncogene.

Our review of the data has shown that recurrent *NFE2L2* mutations were found in multiple cancer types and were found to be associated with poor outcome in head and neck squamous cell carcinoma subjects. Recurrent mutations were identified in: 14.0% of squamous cell lung carcinoma, 13.0% cervical squamous cell carcinoma, 5.5% of infiltrating bladder urothelial carcinoma, 5.3% of endometrial endometrioid adenocarcinoma, 4.2% of head and neck squamous cell carcinoma, and 2.7% of hepatocellular carcinoma. We also investigated *KEAP1*, a repressor of NFE2L2 activity. Mutations in *KEAP1* tended to localize within the NFE2L2 binding domains and did not co-occur with *NFE2L2* recurrent mutations. *KEAP1* mutations were identified in 13.7% of lung adenocarcinoma, 12.4% of squamous cell lung carcinoma, 5.9% of gastric intestinal type adenocarcinoma, and 4.5% of hepatocellular carcinoma. Genes up-regulated in samples from patients with mutations in either *NFE2L2* or *KEAP1* in squamous cell lung carcinoma included many direct NFE2L2 target genes (e.g. GCLC, GCLM, NQO1). Genes up-regulated in *NFE2L2* or *KEAP1* mutant patients significantly associated with genes up-regulated in cell lines resistant to chemotherapy. Using Oncomine® Knowledgebase curated data and cell line exome data we were able to identify cell lines that were representative of populations containing these mutations.

We have provided proof of concept concerning the identification identification of a potentially relevant candidate driver gene using a novel systematic analysis of cancer genomic data.

MATERIALS AND METHODS

To rapidly identify novel genes potentially involved in cancer and new potential drug targets, we developed a systematic method to rank genes through a survey of the compendium of genomic cancer data in the Oncomine® Knowledgebase. A set of gold standard oncogenes and tumor suppressors, as well as genes with no known role in cancer were used to develop an algorithm that evaluated genomic aberrations (including but not limited to focal amplification, recurrent mutations, deleterious mutations, gene fusions, and over-expression) in each gene using samples in the entire Oncomine® Knowledgebase. Numerical rankings were assigned to each aberration type and quantification of the overall results for each target provided the objective genomics-based ranking and hypothesis of functional relevance (oncogene versus tumor suppressor). Once the algorithm was developed, we applied it to and ranked a number of genes with a suspected role in cancer. NFE2L2 was identified; genomic aberrations in NFE2L2 were assessed across the suite of Oncomine® tools and reporter here. Additional details are provided in the figure legends.

CONCLUSIONS

Here we provide a methodology to leverage all available genomic profiling data in the Oncomine® Knowledgebase to rapidly identify genes with potential driver status. Using this methodology, we show that NFE2L2 is a candidate oncogene and provide specific Oncomine® data supporting this conclusion. The gene ranking system used here to identify NFE2L2 as a candidate oncogene can be used to quickly interrogate the world's largest collection of curated cancer genomics data to assess potential cancer relevance of any gene.

RESULTS

Table 1 and Figure 1. Cancer patient data in the Oncomine® Knowledgebase, the world's largest collection of expertly curated cancer genomics data, ranked NFE2L2 as a potential oncogene

Associated Oncomine® Software Application	Tumor Tissue Types	Subject Count
NGS Mutation Browser	66	7,153
NGS Fusion Browser	23	6,438
NGS Expression Browser	23	6,438
NGS Integrative Analysis Browser	74	4,451*
DNA Copy Number Browser	106	28,283
Concepts Edition	632	82,671
Gene Browser	214	31,750

The table on the left indicates patient counts and data types present in the suite of Oncomine® Software as of March 15, 2014. *Subject count for the NGS Integrative Analysis Browser only includes subjects with a full complement of Oncomine® curated genomic data (copy number, expression, mutation, fusion). In the figure below, ranking of NFE2L2 was derived from an analysis of a set of criteria defined across all data in the Oncomine® Knowledgebase. The sum of numeric values across all criteria resulted in NFE2L2 being ranked as a potential oncogene.

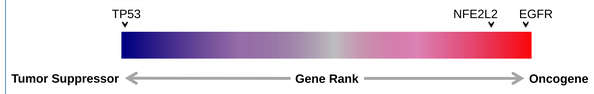
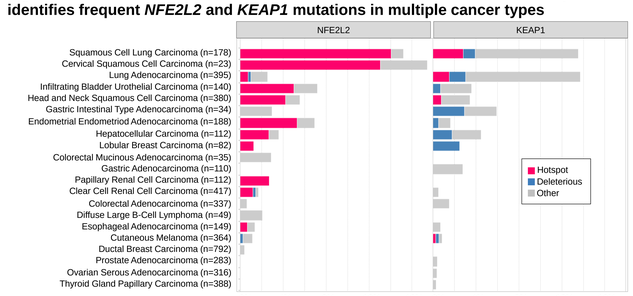
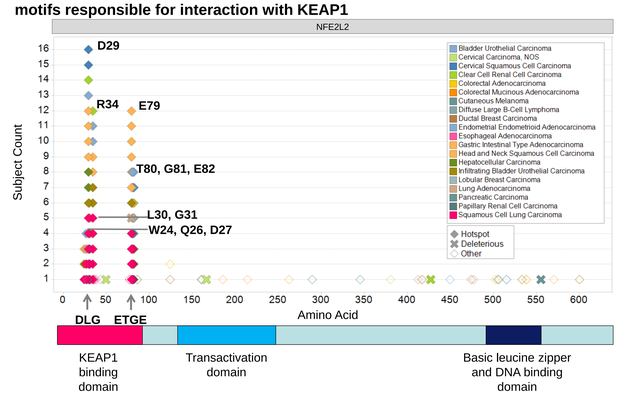


Figure 2. Oncomine® NGS Mutation Browser survey of data from 7,153 patients identifies frequent *NFE2L2* and *KEAP1* mutations in multiple cancer types



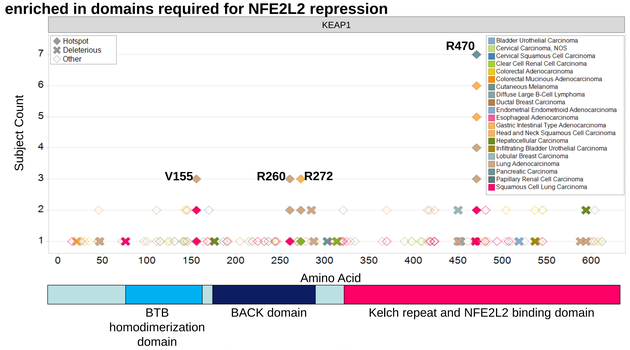
Summary of mutation frequencies for *NFE2L2 and KEAP1* from an analysis of 7,153 patients across 66 cancer types using the Oncomine® NGS Mutation Browser (Q4, 2013 data release). Cancer types are rank ordered by cumulative mutation frequency. Only cancer types with data from more than 20 patient samples and a mutation in either *NFE2L2* or *KEAP1* are shown. Oncomine® Knowledgebase mutation classifications were as follows: hotspot (non-synonymous recurrent mutations in at least 3 patients at the same amino acid position), deleterious (nonsense or frame shift mutation), or other (mutation not defined as hotspot or deleterious).

Figure 3. NFE2L2 recurrent mutations predominantly mapped to the DLG and ETGE motifs responsible for interaction with KEAP1



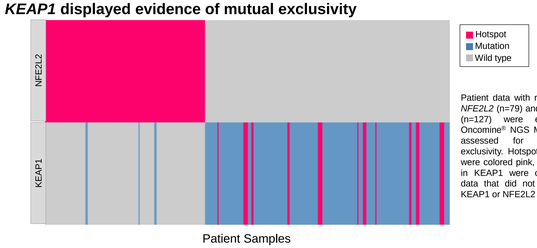
Mutation plots were exported from the Oncomine® NGS Mutation Browser (Q4, 2013 data release). Mutations are colored by individual cancer type and shaped according to Oncomine® Knowledgebase mutation classification. Recurrent mutation sites (>3 mutations at the same amino acid) are indicated. Domain annotations adapted from Mitsuishi et al The Keap1-Nrf2 system in cancers: stress response and anabolic metabolism. Front Oncol, 2012.

Figure 4. KEAP1 mutations and domain map identifies a number of mutations enriched in domains required for NFE2L2 repression



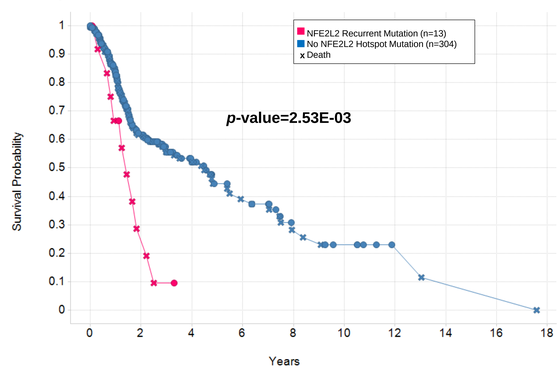
Mutation plots were exported from the Oncomine® NGS Mutation Browser (Q4, 2013 data release). Mutations are colored by individual cancer type and shaped according to Oncomine® Knowledgebase mutation classification. KEAP1 homodimers bind to NFE2L2 to mediate repression. Recurrent mutation sites (>3 mutations at the same amino acid) are indicated. *KEAP1* mutations were enriched in both the KEAP1 homodimerization domain and the NFE2L2 binding domain. Domain annotations adapted from NCBI (NP_036421.2) and UniprotKB (Q14145).

Figure 5. Data from patients with recurrent mutations in *NFE2L2* or mutations in *KEAP1* displayed evidence of mutual exclusivity



Patient data with recurrent mutations in *NFE2L2* (n=79) and mutations in *KEAP1* (n=127) were exported from the Oncomine® NGS Mutation Browser and assessed for co-occurrence and exclusivity. Hotspot missense mutations were colored pink, while other mutations in KEAP1 were colored blue. Patient data that did not have a mutation in KEAP1 or NFE2L2 were colored gray.

Figure 6. *NFE2L2* recurrent mutations were associated with poor outcome in head and neck squamous cell carcinoma



Kaplan-Meier survival curve and analysis taken from the Oncomine® NGS Integrative Analysis Browser (Q4, 2013 data release). *p*-values were generated using the log rank test.

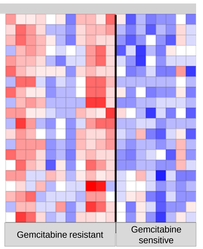
Table 2. An NFE2L2 activation signature included direct NFE2L2 transcriptional target genes

Gene Symbol	*p*-Value	Adjusted *p*-Value
ME1	2.57E-16	1.62E-11
NQO1	9.63E-15	3.03E-10
GPX2	7.31E-11	9.18E-08
GCLM	1.52E-09	1.06E-06
GCLC	2.08E-07	6.15E-05

Differential gene expression analysis was performed on squamous cell lung carcinoma (n=154) patient data from TCGA. Gene lists were ranked by *p*-value according to Student's two class t-test. The adjusted *p*-value was calculated as (*p*-value/*p*-value rank)*number of genes measured. The top1% of genes (n=204) over-expressed in NFE2L2 mutant or KEAP1 mutant squamous cell lung carcinoma were taken as the NFE2L2 activation signature. All signature genes had an adjusted *p*-value<0.005. Examples shown are direct NFE2L2 target genes.

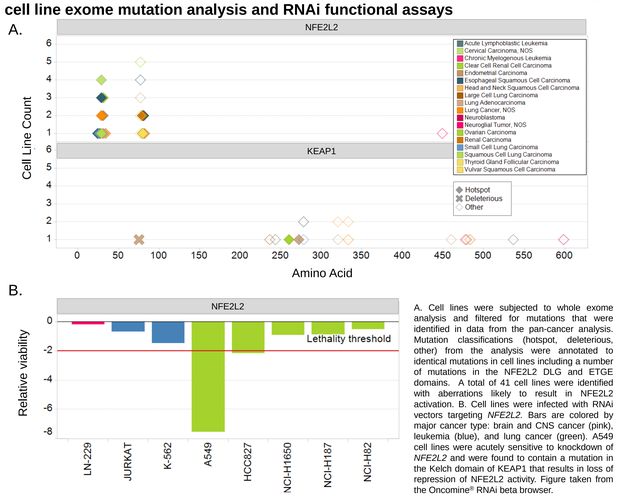
Figure 7. The NFE2L2 activation signature associates with chemotherapy resistance in cell line models

Oncomine® Concepts Edition Property	Signature Overlap	*p*-Value	Q-Value	Odds Ratio
Gemcitabine Resistance	53	1.22E-16	1.52E-13	4.9
Thioguanine Resistance	16	2.47E-12	1.90E-09	13
Cytarabine Resistance	46	4.25E-12	3.16E-09	3.9
Irinotecan Resistance	47	9.93E-12	6.96E-09	3.7
Fluorouracil Resistance	43	2.37E-10	1.27E-07	3.6
Doxorubicin Resistance	42	8.49E-10	4.03E-07	3.5
Vorinostat Resistance	42	8.49E-10	4.01E-07	3.5
Doxorubicin Resistance	39	3.21E-08	1.13E-05	3.1
Methotrexate Resistance	39	3.21E-08	1.11E-05	3.1
Vinblastine Resistance	39	3.21E-08	1.11E-05	3.1



Oncomine® Concepts Edition analysis was performed on the NFE2L2 activation signature (see table above). NFE2L2 activation signature genes were significantly enriched in a number of chemotherapy resistance concepts. The p-value and odds ratio were generated using Fisher's exact test on the association of the NFE2L2 activation signature with the indicated Oncomine® Concept. An example gene association heat map is shown on the right. NFE2L2 activation signature genes were up-regulated in gemcitabine resistant relative to gemcitabine sensitive lung cancer cell lines. Genes are rank ordered by the *p*-value generated from a differential expression analysis. Analysis and figure taken from the Gemma Cell Line dataset in Oncomine® Concepts Edition (Q4, 2013 data release).

Figure 8. Preclinical models of NFE2L2 pathway activation were identified using cell line exome mutation analysis and RNAi functional assays



A. Cell lines were subjected to whole exome analysis and filtered for mutations that were identified in data from the pan-cancer analysis. Mutation classifications (hotspot, deleterious, other) from the analysis were annotated to identical mutations in cell lines including a number of mutations in the NFE2L2 DLG and ETGE domains. A total of 41 cell lines were identified with aberrations likely to result in NFE2L2 activation. B. Cell lines were infected with RNAi vectors targeting *NFE2L2*. Bars are colored by major cancer type: brain and CNS cancer (pink), leukemia (blue), and lung cancer (green). A549 cell lines were acutely sensitive to knockdown of *NFE2L2* and were found to contain a mutation in the Kelch domain of KEAP1 that results in loss of repression of NFE2L2 activity. Figure taken from the Oncomine® RNAi beta browser.

TRADEMARKS/LICENSING