Drug sensitivities in the context of genomic aberrations:

Applications to cancer

Sam Higgins

Oregon Health & Science University

Division of Bioinformatics & Computational Biology

Department of Medical Informatics & Clinical Epidemiology

September 2013

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Thesis of

Samuel M. Higgins

"Drug sensitivities in the context of genomic aberrations: Applications to cancer"

Has been approved

_____

Thesis Advisor

_____

Committee Member

_____

Committee Member

_____

Committee Member

Introduction

Two of the major obstacles that have persisted in the fight against cancer are its heterogeneity and its ability to adapt. Cancer of a given organ system can have a number of different variants, each defined by the unique patterns of cellular aberration driving the cancer's oncogenicity (Koboldt et al. 2012; McLendon et al. 2008; Hammerman et al. 2012). This heterogeneity presents hurdles both in determining the molecular mechanisms driving the cancer and in finding effective treatments for the cancer, as many of the variants tend to respond differently to different treatments. When effective treatments are found, the other obstacle often surfaces -- cancer is often able to adapt and become refractory to treatments, rendering useless many treatments that initially provided a promising outcome (Yonesaka et al. 2011). Both of these obstacles point to the need for new and more advanced cancer treatments that are specifically tailored for an individual patient's cancer strain.

Research consortiums such as the The Cancer Genome Atlas project (TCGA) have made considerable progress using integrative analysis techniques to reveal patterns of cellular aberration associated with a number of different cancers (Koboldt et al. 2012; McLendon et al. 2008; Hammerman et al. 2012; Network 2012) (Koboldt et al. 2012; Hammerman et al. 2012; McLendon et al. 2008; Verhaak et al. 2013). The patterns of cellular aberration identified by these techniques can be used to infer cellular pathways that are likely to play significant roles in driving the cancer being studied. These integrative techniques combine data from multiple global, or *omic* measurements, such as whole exome sequencing or full transcriptome analysis, to provide comprehensive views of the cancer cell's state.

Although the integrative analysis approach has enabled critical steps in revealing the aberrational basis of many cancers, these types of studies do not directly address two critical questions in the fight against cancer. First, they leave open questions surrounding etiology. There are arguments to support that aberrational cellular pathways revealed by these studies do include those that drive or support oncogenicity, as many are found to have aberrations which are believed to facilitate emergent phenomena such as cell proliferation. However, it is difficult to demonstrate, through these approaches, *how* the highlighted cellular pathways functionally affect a cancer. Second, these studies, which integrate multiple aberration data types, are not designed to address the question of how to treat the cancer variants they reveal, although their findings can provide critical information for drug development efforts.

An emerging strategy for cancer treatment is to design what are known as *molecular targeted therapies*. These therapies target the particular cellular pathways upon which a given cancer is likely to be dependent. Where classic chemotherapeutic drugs take the rudimentary approach of inhibiting general cell proliferation, molecular targeted therapies provide the nuance of tailoring treatment so that it can

target elements that are uniquely critical to a particular cancer's survival (B J Druker et al. 1996). These therapies have had major successes, as with the case of imatinib which targets the BCR-ABL fusion protein in chronic myelogenous leukemia (B J Druker et al. 2001), and tamoxifen which inhibits binding of estrogen to estrogen receptors in ER-positive breast cancer (Jordan 2003; Early Breast Cancer Trialists' Collaborative Group 1998).

Assuming suitable targeted drugs exist for a given patient's cancer, the primary challenge is in determining which of the thousands (Wishart et al. 2006) of available drugs will effectively and selectively target a particular patient's strain of cancer. Urgent time frames and negative drug side effects make it critical to administer only those therapies that have a high probability of success.

High-throughput RNA interference screens (RNAi) have provided valuable utility in elucidating useful drug targets in cancer cells (Echeverri and Perrimon 2006). This technique harnesses the naturally occurring phenomena of RNA interference, where short strands of interfering RNA repress the expression of specific gene transcripts. RNA interference can be used to mimic the inhibitory effect of a targeted drug, and thus has been used to systematically probe cancer cells to find their vulnerabilities to down-regulation of particular, potential driver genes (Cheung et al. 2011). Although this technology has aided critical steps in functional genomics and drug discovery, its use is limited by problems such as off-target effects, varying dose efficiency and degradation of the RNAi (Iorns et al. 2007). As well, although RNAi can specifically target and inhibit genes, to the best of our knowledge, no RNAi-based therapy has yet been approved for use in cancer patients.

The *drug screen panels* developed by Tyner et al. offer a new technique to simultaneously screen over a hundred molecular targeted drugs for their effect on a cancer cell line and reveal specific genes that can serve as sensitive drug targets (Tyner et al. 2012; Kulesz-Martin et al. 2013). In doing this, these panels provide an opportunity, similar to RNAi screens, to systematically examine the functional behavior of a cancer. But, unlike RNAi screens, the drug screen panels directly evaluate molecular targeted drugs; many of which are either FDA approved treatments or are already in clinical trials.

There are, however, several hurdles in the application of these panels. Currently, they are capable of testing between 100 and 200 targeted drugs, in turn targeting between 200 and 400 genes. With this number of genes, it is critical to select targeted drugs that will test the cellular elements most likely to play significant roles in the cancer examined. As well, many of the drugs currently available target not one, but a small set of gene products. As a result, the sensitivity of particular genes cannot always be determined.

Individual aberrational genes may not be targetable, or may not provide useful drug targets. However, given that they act in biochemical pathways, which through the collective input of their constituents fulfill particular functions, it is likely there exists effective targets along their respective pathways. Thus, an approach to finding

drug-sensitive gene targets is to construct drug-screen panels that target genes in cellular pathways that show evidence of dysregulation in a given cancer. Here we present a tool that facilitates drug panel design suited to the heterogeneity and vulnerabilities of a given cancer, and exploration of the relationship between genomic aberrations and patterns of drug sensitivity.

Aims

To construct this tool, two aims were achieved.

1) We sought to develop a workflow for using drug sensitivity data to reveal sets of drug sensitive pathways.
2) We sought to develop an evaluation framework for analyzing aberration data types to reveal cellular pathways that are likely to be critical to the survival of a cancer.

Combining the products of these two aims provides a set of scored, prioritized pathways and an integrated view which can help inform a number of decisions in the development and application of drug screen panels and targeted therapies.

This tool promises to be useful to address the obstacles of cancer heterogeneity and adaptation in three ways. The first addresses the risk that a given panel design might leave important driver pathways in *the dark*--untargeted by the drugs on the panel--and thus miss critically important sensitive drug targets. This tool provides an important utility for the construction of drug screen panels. Pathways revealed as critical that cannot be targeted by the currently available drugs might be fruitful areas of investigation for future drug development efforts.

The second is, in providing an integrated view of drug sensitivities and genomic aberrations, this tool can aid investigations into the biological relationship between genomic aberrations and drug sensitivities.

The third is, in highlighting pathways that are both significantly dysregulated and which contain sensitive targets for precision drugs the tool can help inform decisions about the clinical application of targeted therapies.

Methods

Program Design

The tool was constructed as a computer program, which accepts as inputs aberration data, such as somatic mutations copy number alterations and drug-screen data. From these inputs the program produces summaries describing pathway aberration and sensitivity patterns, as well as the overlap between drug targeted and aberrational pathways.

Overall design goals

The program was constructed to meet six overall specifications:

1) Analyze data from either individual patients or a cohort of multiple patients.
2) Determine likely driver pathways from aberration data.
3) Describe the overlap between aberrational pathways and drug targeted or drug sensitive pathways.
4) Allow visualization of aberrations and drug targeting in network diagrams of cellular pathways.
5) Assess performance of peripheral algorithms performing roles such as pathway significance testing and genomic aberration data filtering.
6) Provide sufficient flexibility to allow utilization of alternate pathway repository, gene identification systems, and input aberration data types.

Program architecture

The program was constructed with the R statistical programming language (http://www.r-project.org ; v3.0.1), and uses a set of packages from the Bioconductor code repository(Gentleman et al. 2004).
A schematic representation of the program's architecture can be found in figure 1. The general architecture of the program is to allow input of a given data type (drug screen, genomic, etc.), provide data-type specific summaries, analysis, treatments and filtering, and then restructure the data so that it adheres to a general format, a "patient gene matrix" (PGM). PGMs are constructed to describe the status of each gene in each patient in the cohort, and makes this information amenable as input to a generalized pathway-analysis function. It is noted that, although the program is currently designed to analyze each gene as being in either an "on" or "off" state, original data values for each patient are retained so that enrichment techniques utilizing continuous or multi-leveled data can be implemented in the future. If a single patient has more than one aberration in the same gene, for example more than one damaging variant in the same protein-coding region, the gene will simply be considered aberrational, thus data on repeated mutations will not be utilized.

The generalized pathway analysis function provides two chief roles. First, the function provides general summary statistics describing how many of the genes from the current analysis are annotated to the current set of pathways. These summary statistics include coverage of the pathways by the current platform and

coverage of the genes from the analysis platform by the pathway repository selected.

Second, the summary function provides summary statistics describing enrichment of that pathway in genes affected under a given data type. These statistics include the frequency that the pathway is affected across the cohort, the number of genes in the pathway that are affected, and a selection of statistical tests to assess how significantly each pathway is effected. The default statistical tests currently implemented are the hypergeometric test and the Model-Based Gene Set Enrichment test (MGSA)(Bauer, Robinson, and Gagneur 2011), however, the program was designed so that other pathway analysis test or techniques could be easily integrated.

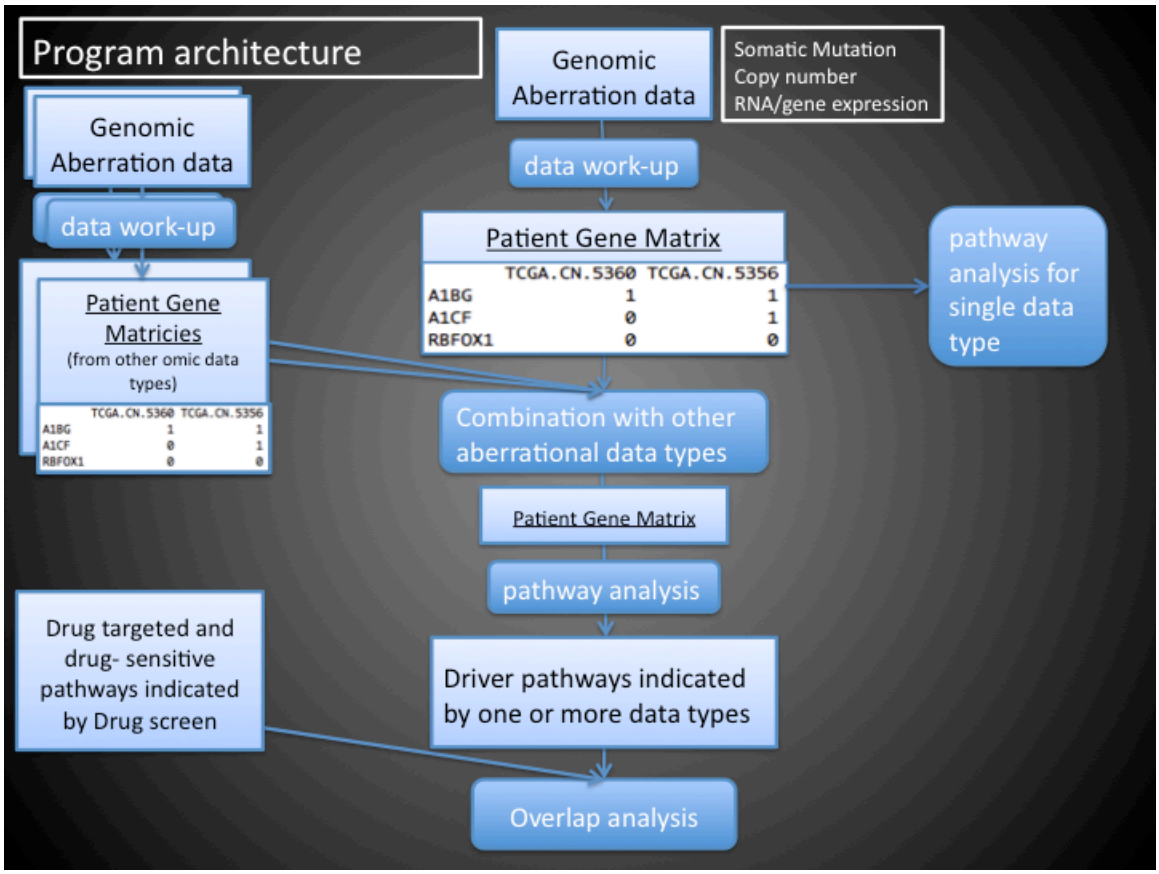


Figure 1: Program architecture. One or more genomic data sets can be input and examined individually for their relation to cellular pathways. If more than one data set is entered, they are combined. Pathway analysis is conducted and the set of putative driver pathways is compared to the set of pathways containing and enriched in drug sensitive targets.

Pathway repositories

Currently, the program allows selection of Reactome, KEGG, NCI, or Biocarta pathway repositories, which are provided via the Graphite Bioconductor package(Sales et al. 2012). Additionally, the program facilitates input of alternate, user provided and/or defined pathway sets. These alternate pathway sets can be provided in the BioPax format(Demir et al. 2010), GSEA format, or bipartite graph

matrix format (using columns as genes and rows as paths). At this time, BioPax is the only available format that will allow for visualization of gene network diagrams by this program.

Gene symbol standards

The program's current implementation requires that genes described in all data types and in cellular pathways be annotated using HGNC/HUGO symbols. Although gene symbol repositories such as Uniprot can provide more stable mapping of symbols, this decision was made because all of the data sources currently utilized by this program employ HUGO symbols. A small script is built into the program that allows the user to check and correct HUGO symbols and synonyms not matching those currently approved by HGNC, then coordinate these symbol corrections so that the same set of symbol correction mappings is used in all parts of the program. In the majority of the analyses presented here, not all gene symbols could be corrected to match approved HUGO symbols.

Visualization

Aberration and drug targeting in pathways can be visualized with use of the *Cytoscape* network visualization program (Shannon et al. 2003; Cline et al. 2007). A script within the program communicates with Cytoscape using the Bioconductor packages, *Graphite* and *Rcytoscape* and the Cytoscape plug-in, *CytoscapeRPC*.

Data processing

Statistical techniques

As a default, the program implements the hypergeometric test to determine significant pathways, and, in turn, highlight them as possible driver pathways. As implemented here, this test examines whether the proportion of genes aberrational or drug sensitive in an individual pathway is significantly higher than the proportion of genes found to be aberrational or drug sensitive across all pathways. This test thus requires that all genes are considered to be either *on* or *off,* with the *on* state corresponding to genes that are considered aberrational or drug-sensitive. As a result of this requirement, continuous or multi-level values must be thresholded or filtered so that genes can be considered "on" or "off" (aberrational/drug sensitive or normal/drug insensitive).

Gene variant data

Two programming modules for processing small nucleotide variant data were constructed. One processes targeted sequence capture data annotated by a custom programmatic pipeline, constructed at Oregon Health & Science University (OHSU). The other script processes somatic sequencing data provided in the .maf format, as specified by TCGA ("Mutation Annotation Format (MAF) Specification - TCGA - National Cancer Institute - Confluence Wiki" 2013). Currently, the program is constructed to ignore variants not annotated with gene identifiers, except for the purpose of providing overall summaries of per-patient variation counts.

The general schema in both modules is to first check the input data for duplicate records and other inconsistencies, then allow filtering based on variant classification, and annotation of any variants with dbSNP database values (Sherry et al. 1999). Individual variants processed by the OHSU pipeline are annotated with one or more sequence ontology terms, as described by the Sequence Ontology project (http://www.sequenceontology.org/index.html or http://uswest.ensembl.org/info/genome/variation/predicted_data.html). Variant annotations described in somatic mutation data are annotated with single variant identifiers as described in the TCGA .maf specification.

The PolyPhen-2 classification algorithm was applied to both the somatic mutation and OHSU variant data to further stratify missense variants by the probability that they will alter their associated protein's value (Adzhubei et al. 2010).

Drug screen data processing

A module of this program was constructed to accept drug screen data. For coverage analysis, this module reads in a set of drug-targeted genes and determines pathways containing targets and the distribution of percent and number of genes targeted in all pathways. For analysis of sensitive targets, this module was constructed with the expectation that drug-screen gene scores, as described in Tyner et al. 2012, will be provided. The program provides the user with a distribution of drug screen scores and allows the user to select, with the distribution as a visual aid, a cutoff value to differentiate sensitive and non-sensitive targets.

Single-patient analysis versus cohort analysis

The design requirements of this program necessitate that the program be capable of analyzing either single patients or cohorts of patients. When the program is being used to analyze data from an individual patient, the set of genes found aberrational or drug sensitive in that patient can be used directly in the hypergeometric test. However, when the program is used to analyze cohorts of patients, the data from the cohort must be "collapsed" into a single set of genes that are considered to be either *on* or *off*. To carry out this collapse, the default option for the program is to consider a gene to be in an *on* state if it is found to be *on* in any member of the cohort. However, two other options are available, allowing a gene to be considered *on* if it is found *on* in more than a user-defined number or proportion of patients in the cohort.

Reduced coverage analysis

When data is provided from a low-coverage platform, such as the drug screen, which analyzes 500 targeted genes , the full set of genes in the chosen pathway repository generally cannot be used. To account for these limited-coverage situations, before pathway significance calculations are performed pathways from the repository are limited to include only those genes analyzed by the analysis platform.

Simultaneous analysis of multiple aberration data types

The program allows multiple aberration data types, such as copy number and RNA sequencing, to be analyzed and compared. The technique currently implemented for combining multiple data types is to first classify genes as aberrational or normal, with each data type independently, then merge sets of aberrational and normal genes together, weighting each aberration type equally.

Data procurement

TCGA

Publically available somatic mutation data sets for 75 acute meyloid leukemia (AML) patients and 323 head and neck squamous cell carcinoma (HNSCC) patients were downloaded from The Cancer Genome Atlas project data portal website on June 20th, 2012 and April 18th, 2012, respectively.

AML data from OHSU

Targeted sequence capture data was obtained for 14 of the 34 AML patients analyzed in Tyner et al, 2012. This sequencing data was produced for a study examining patterns of aberration in the tyrosine kinome, thus the coverage of the genes targeted is biased toward kinases, phosphatases, and other kinase-relate genes. For a more detailed description of how this data set was produced, see Loriaux et al. (2008).

Drug screen data

The drug screen data used in the analyses presented here comes from two sources: Tyner et al. (2012) and the lab of Dr. Molly Kulesz-Martin at OHSU. The Tyner et al

drug screen data was produced by assaying cancer samples from 151 leukemia patients against 66 different targeted therapies, and is publically available through the supplementary material available on the journal's website. The Kulesz-Martin drug screen was designed for the assay HNSCC patients. Only the target spectrum from this drug screen panel is analyzed here.

Results

Three use cases were explored to demonstrate the utility of this tool. The first use case examines the overlap of drug-sensitive pathways and putative driver pathways found when examining drug screen data and sequence variant data both produced from the same cohort of 14 AML patients. The second use case examines the same overlap, but utilizes drug screen data from 34 AML patients and a publically available, high-quality sequence variant data set from a much larger cohort of 75 AML patients.

This demonstrates the interaction between aberrations and drug sensitivities in the context of cellular pathways. Next this explores how a broader survey of aberrational pathways can be used to get a better picture of the range of likely driver pathways and thus assess how well the panel design actually targets critical driver pathways. Together, these first two use cases compare the use of aberration data and drug screen data from the same patients with the use of aberration and drug screen data from separate patient cohorts.

One caveat of this comparison is that DNA variants from use case one and two are determined in different manners. In use case one, variants are determined by comparing the DNA sequence from a tumor sample to the corresponding sequence in the human reference genome. On the other hand, the variants in use case two were determined by comparing the sequences found in tumor and normal samples from the same patients. The results from these two approaches will differ in several ways. The unpaired data should show a much higher per-patient mutation rate, because the variants found reflect the genetic differences expected between unrelated individuals, as opposed to the variants that arise in the life of an individual. A critical technique used to ameliorate the large number of variants found in unpaired sequencing data is to filter out variants found in the dbSNP database. dbSNP is a database of variants which are commonly found across the population(Sherry, Ward, and Sirotkin 1999). Because of their frequency in the population, variants found in dbSNP are regarded to be generally unassociated with disease processes. However, it cannot be strictly assumed that they will never contribute to diseases and it is entirely possible some disease-associated variants are found in dbSNP and were ignored in the analysis because of this. A corollary issue is that because use-case two only examines variants that arose in the life of the patient, any disease causing or supporting variants that are found in the matched-

normal sample and the tumor sample, (ex: a disease causing gene inherited by a patient from their parents), will be ignored.

The third use case explores the utility this program can offer to the design of panels being built to examine HNSCC. A large, high quality sequence variant data set is used to determine likely driver pathways. This set of driver pathways is then compared to those targeted by a drug screen panel to determine which pathways are left *dark* by it's set of targeted therapies.

Use case 1: drug screen and sequence capture data from same cohort

Drug screen data description and results

In all analyses of drug screen data presented here, the gene scores described in Tyner et al, 2012 are utilized to determine drug sensitivity of genes and pathways. In all use cases a gene score cutoff of 40 was used to differentiate between sensitive and insensitive gene targets. This cutoff score was empirically determined by visual inspection of the distribution of all gene scores for the full cohort of 34 AML patients from Tyner et al, 2012 (see figure 2). After applying this cutoff, only 10 of the 14 matched patients were found to have genes sensitive to the drugs on the panel. As the goal of the analysis presented in this use case was to examine how patterns of genomic aberration directly relate to patterns of drug sensitivity, the 4 patients showing no drug sensitivity were excluded from the overall analysis, reducing the cohort in both the drug screen and the sequence capture to 10 patients. A comparison of the gene scores for the 4 patients filtered out and the 10 remaining can be found in figure 2.

Among the cohort of 10 patients, 49 genes were found to be drug-sensitive targets, with individual patient samples having between 1 and 26 sensitive targets. Of the drug sensitive genes, 31 were found to have annotation to pathways in Reactome, indicating 244 pathways as containing drug sensitive targets.

A pathway enrichment analysis of drug-sensitive targets revealed 104 pathways to have significant enrichment in drug-sensitive targets(hypergeometric p-values< 0.05; false-discovery rate adjusted). However, of the 104 enriched pathways, 37 had only a single drug-targeted gene. A table of pathways containing drug-sensitive genes can be found in the use case 1 supplement.

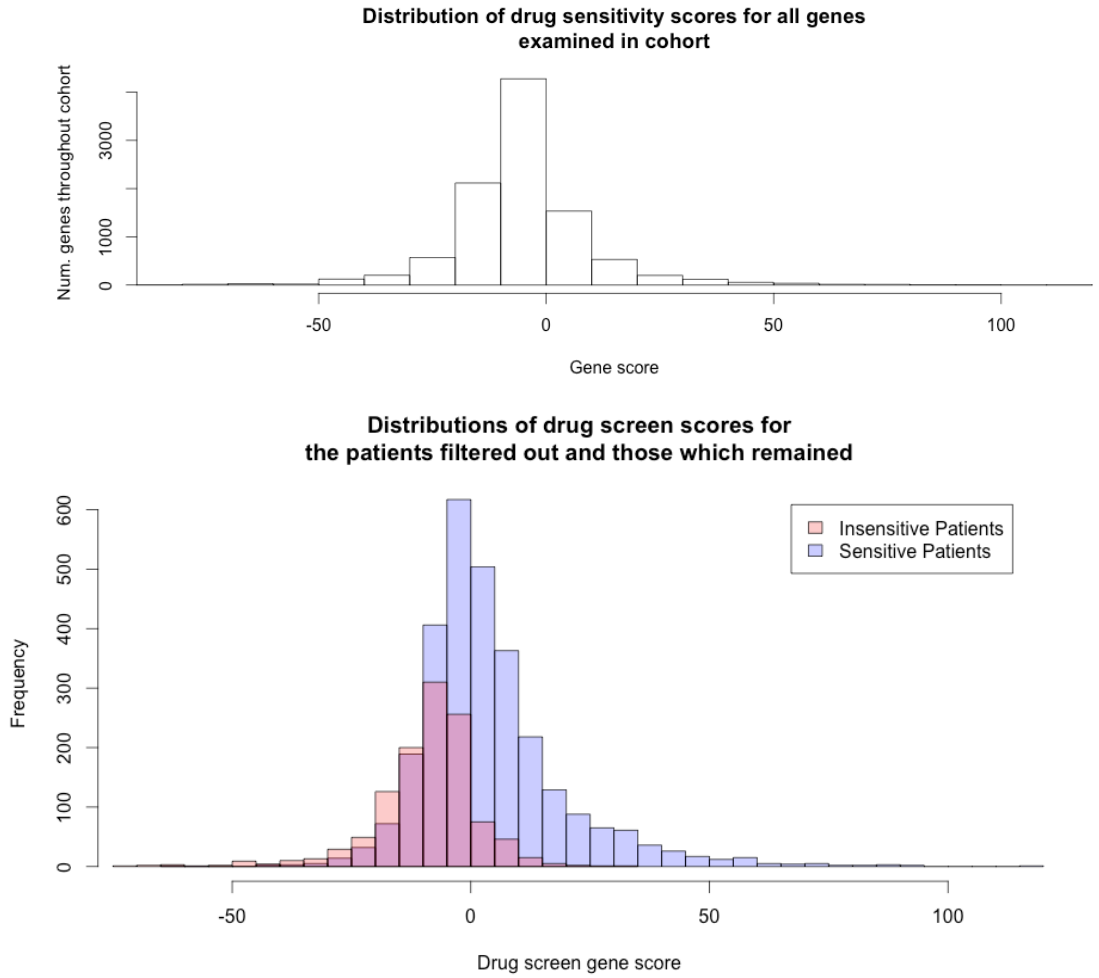


Figure 2. Top: distribution of drug sensitivity scores in 34 AML patients. Bottom: comparison of drug screen scores in the 4 patients that were found to have no sensitive gene targets and were excluded from the cohort to those scores in the 10 patients found to have sensitive gene targets.

Sequencing data

The set of genes targeted in the targeted sequence capture data was selected specifically to explore the mutational landscape in the tyrosine kinome of several types of leukemia (Loriaux et al. 2008; Tyner et al. 2008), and thus this data only achieves coverage of a set of 3501 unique genes, among which are a large proportion of kinases, phosphatases and other kinase-related genes. Of the 3501 genes that were sequenced, 1230 had annotation to 984 Reactome pathways. A histogram describing the distribution of pathway coverage can be found in the Appendix. To adjust for the low sequence capture coverage, the set of Reactome pathways used in the enrichment analysis was limited to include only those genes that were sequenced.

In the cohort of 10 patients, 772,404 sequence variants were found annotated with HUGO symbols; of these, 751,857 had records in the dbSNP database and were filtered out, leaving 20,547 variants. Filtering was then done by variant type after annotating missense variants with PolyPhen predictions (Polyphen predictions allowed 206 of 387 PolyPhen-annotated variants to be filtered out). A list of variant types retained and discarded can be found in the Appendix, along with distributions of variant types found in the cohort. Filtering by variant type produced a final set of 236 unique, qualified aberrations, with individual patients having between 9 and 55 aberrations each.

Paths enriched

Pathway enrichment analysis using aberrations from across the cohort showed 69 Reactome pathways to be significantly enriched in aberration. A table of these pathways can be found in the use case 1 supplement.

Overlap analysis

287 genes were both sequenced and targeted by drugs on the panel, however, only 4 of these, "EPHA8", "ERBB2", "FLT3" and "RET" were found, among all members of the cohort, both to be aberrational and to be drug-sensitive targets.
20 pathways were found to contain sensitive targets and to be enriched in aberration; of these, seven were also found to be significantly enriched in drug-sensitive targets. However, only two of these 7 pathways, *CD28 co-stimulation and Semaphorin interactions*, were found to have more than 1 aberrational and 1 sensitive target. A table comparing pathways containing drug sensitive gene targets to their aberrational composition can be found in the use case 1 supplement.

Dark pathways

Examining the cohort as a whole, 22 pathways were found to be "dark", with significant enrichments in aberration, but with no genes targeted by the drugs on the leukemia panel. (Among individual patients, between 1 and 9 pathways were found to be dark.) These pathways contained between 1 and 3 aberrational genes. The only 4 pathways with more than one aberrational gene were also found to have the largest number of testable genes. A table detailing all 22 dark pathways can be found in the use case 1 supplement.

Cryptic pathways

We also examined cryptic pathways: those which show evidence of drug sensitivity, but given the aberrational analysis platforms utilized, do not show evidence of significant levels of aberration. In all, 211 pathways were found to contain sensitive targets but were not found to have significant aberration. Of the 211 pathways, 118 were found to have some aberrational genes.

Use case 2: drug screen data + somatic sequencing data from TCGA

Use case two analyzes the aberrations found in a larger cohort of AML patients to provide a better picture of the spectrum of gene aberration patterns, and, by extension, the spectrum of pathways likely to be associated with AML. To get a

better idea of the potential spectrum of aberrational pathways in AML, and answer the question of how well the panel from use case 1 targeted driver pathways, we analyzed somatic mutation data for a cohort of 75 AML patients.

Drug screen data from all 34 AML patients from Tyner et al, 2012 was analyzed in the second use case. Of the 34 patients, however, only 14 were found to have targets that were sensitive to the drugs on the panel. Additionally, with the 4 additional patients, only 3 additional genes, "FLT1", "FLT4" and "MAPK12", were found to be drug sensitive. As a result, the set of pathways containing drug-sensitive genes was the same as the set found in use case 1.

Somatic mutation data

Where in use case 1, 10 patients were analyzed for aberration using targeted sequence capture, use case 2 analyzes somatic mutation data from a cohort 75 AML patients. In the unfiltered data from this cohort, 1354 unique variants were found, with individual patients having between 1 and 177 variants each. Sequence variant types were annotated differently in the somatic mutation data, although the different annotations refer to many of the same variants. PolyPhen was again used to predict damaging missense mutations, and to further annotate the missense mutations. Figure 3 shows the distribution of variant types found annotated to the top 20 most variant genes, before filtering. The variant types selected to qualify variants as valid aberrations can be found in the Appendix. Filtering by variant type removed 759 unique variants; another 54 were found to have their gene symbol annotated as “Unknown” and were removed, and 17 genes were found to have more than one variant in a single patient. Because somatic mutations are considered to have occurred in the life of a patient, those with dbSNP values have a higher probability of being directly associated with a patient’s cancer. Thus, while variants with dbSNP records were filtered out in the case of the sequence capture data from use case 1, they were not filtered out of the somatic mutation data.

After all filtering, 578 variants remained, and were considered as the set of valid aberrations in this cohort, for the somatic mutation data type. Individual patients were found to have between 1 and 57 aberrations each; a distribution of aberrations

per patient can be found in the appendix.

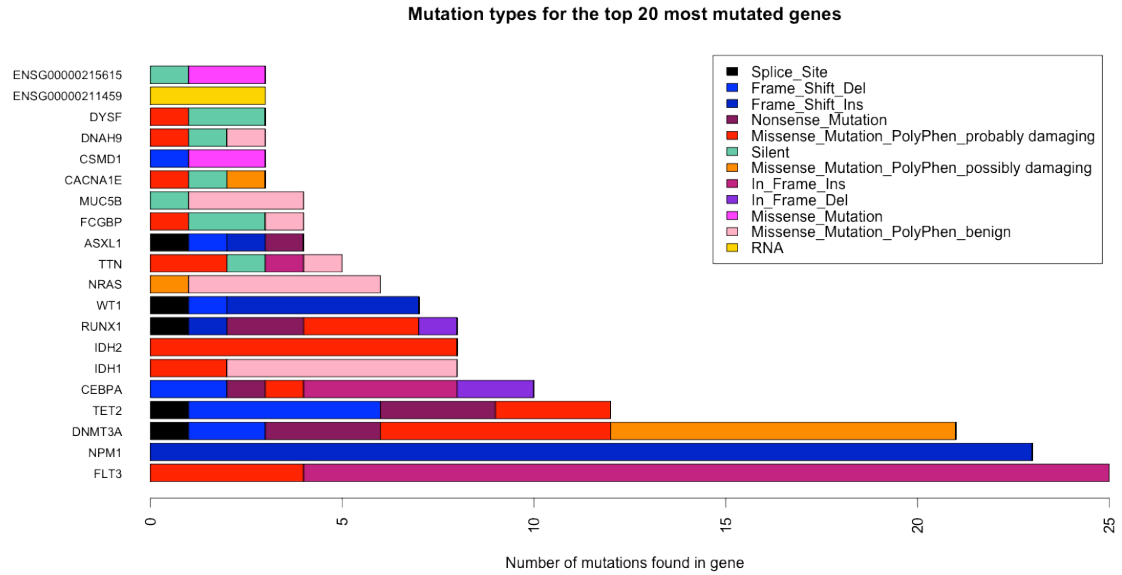


Figure 3: Distributions of variant types in the top 20 most variant genes found in the somatic mutation data of 75 AML patients.

462 unique HUGO symbols were found annotated to the final set of 578 aberrations. Of these symbols, 147 were found annotated to 522 Reactome pathways. Of these 522 pathways containing aberrational genes, pathway enrichment analysis revealed 71 to have statistically significant numbers of aberrational genes, however, 22 of those paths considered enriched contained only one aberrational gene. A table of these significantly enriched paths can be found detailed in the use case 2 supplement.

Overlap analysis

Prioritized pathways

Of the 45 pathways that were enriched in aberration and were drug-targeted, 36 were also found to contain sensitive drug targets. A further 8 of those 36 pathways were also significantly enriched in sensitive targets. Interestingly, all of the 8 paths that were significantly enriched had 2 or more drug sensitive targets; whereas, in use case 1, 6 out of the 7 pathways that were aberration-enriched and sensitive-target enriched had enrichments in sensitive targets that consisted of only one gene.

Dark pathways

This analysis revealed 26 dark pathways that could provide valuable targets for future drug panel designs. Together these 26 pathways contained 163 genes; 29 of which were found in more than one of the aberration-enriched pathways. However, only 10 of the pathways had more than one aberrational gene and the remaining 16 had testable path lengths of 5 or less. A table detailing the enrichments found in the dark pathways can be found in the use case 2 supplementary materials.

Cryptic pathways
Evidence was also found that other types or larger samples of aberration data might reveal other dysregulated pathways, as 208 pathways were found to contain drug-sensitive targets, but not to have significant enrichment in aberrational genes. Of these 208 pathways, 58 showed no aberrational genes at all, and the remaining 150 had aberrational genes, but in quantities that did not achieve statistical significance. 81 of the 208 pathways have significant enrichment in sensitive genes, however 37 of these pathways had a testable path length of 1. A table detailing the cryptic pathways can be found in the use case 2 supplementary materials.

Comparing use case 1 with use case 2:

Parameters of comparison:
Use cases 1 and 2 serve to compare using the same patients for aberration and functional analysis with using different cohorts for aberration and functional analysis. In both cases the same aberration data type was analyzed--DNA sequence variants--and the same set of drug screen data, all from the AML patients presented in Tyner et al, 2012. However, because the DNA sequencing techniques differed, several adjustments were necessary to carry out a more accurate comparison. First, the somatic mutation data was coverage-limited, reducing the data set to include only those genes sequenced by the sequence capture analysis from use case 1. As well, dbSNP mutations were filtered out of the somatic mutation data to better match the processing from the sequence capture.

Gene by gene and pathway by pathway comparison
Comparison of aberrational genes identified by both approaches showed only 4 genes that were found by both sequencing approaches (TTN, FLT3, SMG1 and AP1G2). Pathway enrichment analysis indicated the two approaches had 21 pathways in common that were enriched in aberration, however, only four of these pathways contained more than one aberrational gene.

Four pathways, *Cell death signaling via NRAGE, NRIF and NADE*, *NRAGE signals death through JNK*, *Prolactin receptor signaling and Signaling by constitutively active EGFR,* were found to be enriched in aberration and to contain sensitive targets by both approaches.

Use case 3: Somatic sequencing data and drug screen coverage

The final use case answers the question of how well a particular drug screen panel design targets pathways showing significant enrichment in aberration.

Somatic mutation data from TCGA
Somatic mutation data from a cohort of 323 HNSCC patients was analyzed for pathway enrichment. Of the 67,835 unique variants found across the cohort, 40274 were filtered out as benign, yielding a set of 27,111 qualified aberrations. Variant types selected to imply valid aberrations can be found in the Appendix. 1,553

variants had records in dbSNP, but, because these mutations are somatic, there is a higher probability they are associated with pathogenesis, thus the dbSNP variants were not filtered out. Individual patients had between 13 and 2,238 variants before filtering; of these variants, between 4 and 821 variants were ultimately qualified as valid aberrations (not counting multiple aberrations in the same gene).

Somatic mutation Enrichment

10,820 unique HUGO symbols were annotated to the 27,111 qualified aberrations; of these HUGO symbols, 3,734 could be found annotated to Reactome pathways. Pathway enrichment analysis was then conducted with this set of genes and revealed 133 pathways to be enriched in aberration. A table detailing pathway enrichments can be found in the use case 3 supplements.

Drug screen coverage

The drug screen panel assessed for coverage of aberrational paths contained 129 drugs, together targeting 385 genes. Of these genes, 189 could be found annotated across 537 Reactome pathways. A figure comparing the distributions of drug-targeted genes to the number of drug-targeted genes in pathways can be found in the Appendix.

Dark pathways

Comparing the set of pathways found to be enriched in aberration to the set of drug targeted pathways, a total of 76 dark, aberration-enriched, not drug targeted pathways were found. Together these pathways provide 931 possible new gene targets for drug panel development. The current panel design did target 57 of the aberration-enriched pathways, however, 12 of these pathways contained only single drug targets, including 2 pathways that were found to carry aberrations in more than 20 patients from the cohort. A table detailing the enrichments found in the dark pathways found for the current HNSCC panel design can be found in the use case 3 supplementary materials.

Discussion

Use case 1

Data from the drug screen panel examined in use case 1 indicated that among the 10 patients analyzed, there were 49 drug sensitive gene targets. However, these are likely not the only sensitive targets in this cohort. The panel targeted only 290 genes and, given the drug score cutoff selected, 4 of the cancer samples analyzed showed no genes to be drug sensitive (Of 34 AML patients from Tyner et al. 2012, 20 had no genes showing significant sensitivity). The tool presented here used genomic aberration data taken from the same set of patients to reveal 22 *dark* pathways, who's aberrational nature suggest they play roles driving the cancer, but which are not targeted by any drugs on the panel. This result suggests future panels constructed to target these pathways are likely to find additional sensitive gene targets.

Another 20 cellular pathways were revealed both to contain sensitive targets and to have statistically significant levels of genomic aberration. Two of these pathways, which were of particular interest, are the *Nef and signal transduction* pathway and the *Signaling by constitutively active EGFR* pathway. With three out of three targeted genes showing sensitivity, the *Nef and signal transduction* pathway had the highest proportion of sensitive gene targets. In addition, aberrations in this pathway suggested it might play a role as a driver of cancer, as 1 out of the 4 genes sequenced was found to be aberrational.

The other pathway, *Signaling by constitutively active EGFR*, showed one out of two tested genes to be sensitive drug targets and 1 out of 9 sequenced genes to be aberrational. Because only 1 aberrational gene in each of these pathways gave the pathways statistically significant enrichment in aberration, the reliability of statistical enrichment as a metric to reveal driver pathways is questionable. In cases such as this one, where limited genomic coverage and small sample size affect the reliability of conclusions determined from statistical calculations, the pathways prioritized might be considered as prioritized for further genomic or drug screen analysis, as opposed to clinical use or immediate panel construction options.

A crucial detail of the analysis conducted in use case 1 is that the drug screen data and the genomic aberration data was produced from the same set of patients. Results from analysis of individual patients can be seen in table 1.

Table 1: Overlap of sensitive and aberrational pathways in individual patients.

Patient Number	Pathways aberration-enriched but not targeted	Pathways aberration-enriched, without sensitive targets	Pathways aberration-enriched and drug sensitive	Pathways not aberration-enriched, but drug sensitive
7.00335	6	4	3	103
8.00024	9	10	1	50
8.00053	7	64	3	14
8.00076	1	22	0	14
8.00102	2	2	3	109
9.00256	4	10	0	94
9.00453	2	21	6	58
9.00454	6	44	32	95
9.00473	8	10	0	30
10.00136	7	2	5	150

As can be seen in table 1, overlap is generally seen between drug sensitive and aberrational pathways. Where overlap is not seen in three of the patients, it is likely the genomic aberration analysis or drug screen analysis did not have wide enough coverage to detect genes in those pathways that would have overlapped.

Use case 2

In use case 2 we investigated more deeply and more broadly the patterns of aberration associated with AML by analyzing a higher quality genomic aberration data set from a much larger cohort of AML patients. Comparing the results from this genomic aberration data with the targeting spectra of drug panel, 26 dark pathways were found, which had levels of aberration suggesting they might drive strains of AML, but which were not targeted by drugs on the panel. Three of the pathways from use case 2 were of particular interest and are judged as more promising candidates for drug targeting: *Interaction between L1 and Ankyrins*, *Cohesin Loading onto Chromatin* and *Establishment of Sister Chromatid Cohesion*. These paths are relatively small, containing 26, 10 and 11 genes, respectively, and they are seen repeatedly aberrational in the cohort, with 4, 10 and 10 individuals effected, but the repeated aberration was not found to be due to a single frequently aberrational gene in either of the three cases.

The product of use case 2 also allowed us to evaluate some of the results from use case 1. For instance, *Nef and signal transduction* pathway, which appeared aberration-enriched in use case one, did not show significant enrichment when examining the larger, higher quality data set from use case 2. While it is possible that the *Nef and signal transduction pathway* in fact drives a very rare strain of AML, this result suggests the 1 gene enrichment seen in use case 1 was a spurious enrichment resulting from the background noise of randomly occurring mutation.

In contrast to the Nef pathway, analysis of use case 2's genomic data provided support for the conclusion from use case 1 that the *signaling by constitutively active EGFR* pathway was a likely driver pathway, as this pathway again showed significant enrichment in genomic aberration using the somatic mutation data from use case 2.

The results from use case 2 highlight this tool's utility in revealing both known driver pathways and new pathways not currently associated with AML. The *Signaling to RAS* pathway and the *GRB2 events in ERBB2 signaling pathway* are both well established as being of particular significance to AML. With 34 genes, the *Signaling to RAS* pathway is of a manageably small size yet among the cohort, 7 out of 14 targeted genes were found to be sensitive targets. As well, generally larger numbers of patients see aberration and drug sensitivity in this pathway -- 8 of the drug screen cohort's patients show drug sensitivities along this path, and 5 patients from the somatic mutation cohort show mutations along this path. Finally, RAS is a well-known oncogene, which has previously been reported to play a role in AML(Grossmann et al. 2013; Goodsell 1999) .

The *GRB2 events in ERBB2 signaling* pathway provides another example of a pathway highlighted by the approach presented here, which is supported in literature as being associated with cancer and the specific cancer examined here. In the data analyzed here, this pathway shows significant enrichment for aberration and drug sensitive gene targets. Although references in literature to the particular Reactome pathway name *"GRB2 events in ERBB2 signaling"* are scarce, ERBB2

signaling pathways are widely implicated in cancer (Hynes and MacDonald 2009; Yonesaka et al. 2011) and in AML (Martín-Subero et al. 2001).

Finally, the *Cell surface interactions at the vascular wall* pathway provides an example of a novel pathway revealed by the tool presented here. To the best of our knowledge, this pathway has not been implicated as having a role in AML. However this pathway shows 5 out of 9 targeted genes to be sensitive targets, and 9 genes to be aberrational, the highest number out of the 8 aberrational pathways showing drug sensitivities. This pathway *is* also the largest out of the 8, and its percent of aberrational genes is relatively low, but still statistically significant.

One patient from use case 1 showed aberration in the *Cell surface interactions at the vascular wall* pathway. Although the *Cell surface interactions at the vascular wall* pathway was not found in this patient's set of drug sensitive pathways, this patient's drug screen scores for several of the genes in this pathway were relatively high. Two of them, LCK and YES1 were in fact right at the threshold established for sensitivity, with scores of 39.25 and 39.675. In figure 4, a network diagram is provided which illustrates the relation between the aberrational and drug sensitive genes in this pathway, with the drug screen sensitivity threshold lowered from 40 to 39. The observance of drug sensitive targets and genomic aberrations in this pathway provides support for it playing a significant role in some of the AML strains analyzed here.

Comparing aberrational pathways in use cases 1 and 2

One expectation is that examination of a larger cohort should reveal a greater proportion of the pathways expected to play a role in a cancer. Comparing the significantly aberrational pathways found in use cases 1 and 2 appears to confirm this expectation, as the analysis of the 75 patients yields 109 pathway to be significantly enriched in aberration, while analysis of the sequence capture data highlighted just 69 pathways. However, there were 48 pathways which showed enrichment by the sequence capture data that did not see enrichment using the somatic data. There are a number of reasons why we might expect to find pathways that only appear enriched in the sequence capture cohort. One is that the larger cohort will not necessarily capture the full range of aberrations that can drive the cancer in question, or show significant enrichment in all the pathways associated with those aberrations.

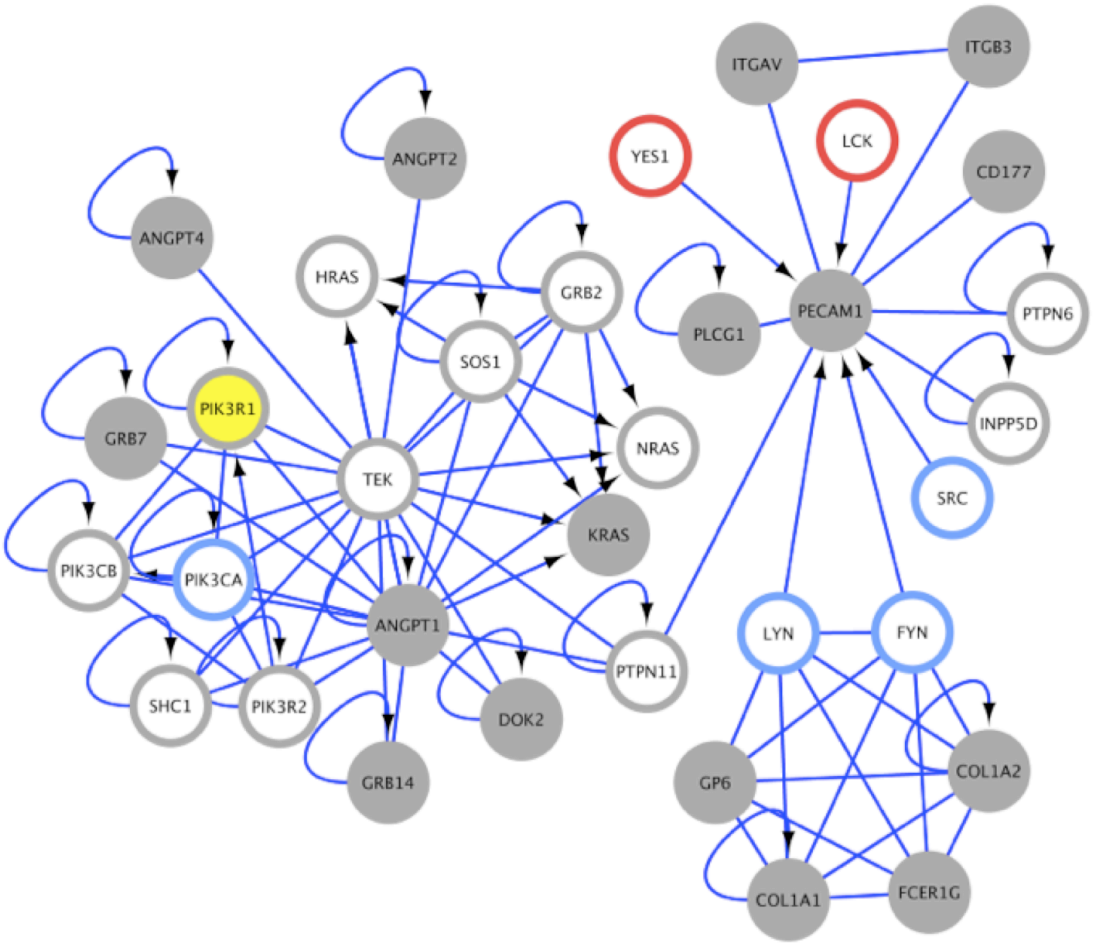


Figure 4: A sub network of the 92-gene pathway, *Cell surface interactions at the vascular wall,* displaying the aberrational genes (yellow), drug sensitive genes (red rings), drug insensitive genes (blue), and genes not included in the sequence capture and drug screen coverage (grey circles and rings, respectively) as found in patient 9.00454

There are also several reasons why direct comparison of these results is likely to be misleading. First and foremost, different sequencing techniques were used to generate the data in use cases 1 and 2. Because the sequencing from use case 1 did not employ an approach to highlight somatic mutations, many of the variants from use case 1 which were qualified to be aberrations are likely just noise—random variants unassociated with the cancer. Another issue is that a low overall mutation rate and small cohort size will lead to a situation where pathways, especially smaller ones, will show statistically significantly enrichment with the presence of just a single aberrational gene. This problem is multiplied by the effects low coverage, because the lengths of many pathways are reduced, and by the greater number of random, unspecific mutations found in variant data sets produced without matched normal tissue samples.

Table 2: Comparison of the aberration data sets used in the three use cases

	Use case 1: AML	Use case 2: AML	Use case 3: HNSCC
Aberration data type	Targeted sequence capture variants	Somatic mutation	Somatic mutation
Number of patients	10	75	323
Mutations per patient, before filtering	78910 to 81730	1 to 177	13 to 2238
Mutations per patient, after filtering	9 to 55	1 to 57	4 to 821
Unique aberrational genes in cohort	59	462	10820
Number of pathways with aberrational genes	208	522	1204
Number of pathways found enriched in aberrational genes	69	71	133

Use case 3

Among the 133 paths enriched for aberration in the HNSCC cohort are 76 *dark* paths, not targeted by the drugs selected for the current panel design.
In this set of 76 paths are 931 genes that might serve as drug screen targets. As the current panel design targets less than 400 genes, careful selection of gene targets is necessary so that the largest and most pertinent set of cellular pathways can be tested for drug sensitivity.

One technique for determining important targets is to select those pathways that are most commonly seen affected across the cohort. A number of summary statistics provided by this program can help to highlight the most interesting of these pathways. The *Ion Channel Transport* and *NCAM1 interactions* pathways are of interest because they are found to be aberrational in 34% and 38% of the cohort. Finally, mutation rates in these pathways are more spread out, with individual genes not mutated in more than 12 patients, but with overall counts of aberration in these pathways across the cohort relatively high (156 and 175 aberrations, respectively). This low hanging fruit approach is important, but given the demands of personalized medicine, which urges treatment of individual patients as opposed to something that approaches the median of a population, only focusing on the most commonly aberrational pathways may be a undesirable approach. Another tactic is to select genes that are annotated to multiple pathways. Using this approach, we see that genes such as ARHGEF9, a GTPase involved in cell signaling, and ITGB1, an

integrin family membrane receptor involved in cancer metastasis (Brakebusch and Fässler 2005, 1), are found in multiple dark pathways (6 and 5, respectively)

Hypergeometric not ideal, though widely used

The hypergeometric test was included as the program's default method for finding possible driver pathways based on gene aberration data. Use of this test here makes the assumption that if a pathway is found relatively enriched for mutation in a cancerous cell, then it likely has a role in driving the cancer. False discovery rate adjustment for multiple testing must be used because of the large number of pathways examined. This adjustment technique assumes pathways are not overlapping; but as there is significant overlap between pathways, the number of pathways found to be significant might be misleading.

Another drawback is that this technique does not leverage the specific topology of a cellular pathway network in determining the significance of gene aberrations or drug targeting, rather, all genes are given the same weight. This is undesirable because it is known that some genes in fact act as hubs, connecting large numbers of other genes, or as critical bottle necks in information flow or rate limiting steps in metabolic pathways. So-called "topology-based" pathways analysis techniques, such as PARADIGM, have gained wide acceptance for their power in elucidating the significance of aberrations in a pathway(Eifert and Powers 2012; Ng et al. 2012; Vaske et al. 2010).

Enrichment for drug sensitive targets

Although statistical enrichment for drug-sensitive targets in a pathway *is* an interesting result, and is useful for prioritizing pathways, for several reasons it may be less of an important metric than enrichment is in the case of aberration analysis. A second reason is that while molecular targeted drugs generally down-regulate the activity of a gene, it is likely not all genes along a driver pathway will have functionality that will aid the oncogenicity of a cancer. Thus, in a pathway that is most ideal for drug targeting, there are likely numerous genes, whose inhibition with targeted drugs would not have a desirable effect on a cancer.
As well, the limited coverage of the drug screen panels can give many pathways significantly significant enrichment even if only a single drug-sensitive target is found.

Conclusion

Here we have developed a tool for evaluating drug targets, drug sensitivities and genomic aberrations in the context of cellular pathways. This provides critical information to aid in the selection of targeted therapies and in the construction of future drug screen panels.

It is centrally important that the drug screen's functional analysis does not leave important pathways in the *dark*—untargeted by the drugs on a panel. The tool presented here provides a framework for *in silico* modeling of the relation between

a drug screen panel's target spectra and the likely driver pathways in a cancer. Thus, this tool provides key information for assuring likely driver pathways are illuminated by the drug screen analysis.

Drug screens can suggest a number of drug targets. In some situations, the degree of a target's sensitivity can make it stand clearly apart from other targets; however, this is not always the case, and additional information might be needed to make informed treatment decisions. Leveraging knowledge of aberrations in particular genes *can* help to highlight genes that can serve as sensitive targets, but this might only be an effective strategy in the case of activating mutations and provides only single targetable genes. Putting drug sensitivities and aberrations in a pathway context allows the expansion of target lists and, conversely, can help to stratify lists of potential targets based on the aberrational status of the cellular pathways in which they operate.

As well as providing a way to put drug sensitivities in the context of aberrations, this tool can be used to place genomic aberrations in a functional context, and help guide selection of analysis platforms for revealing disease-causing cellular dysregulation.

Although the use cases presented here analyze cancer data from drug screen panels, this tool is agnostic of both disease and functional analysis platform. RNAi functional genomic screens and drug screen panels both probe individual genes in pathways to determine points of sensitivity. And, as do the drug screen panels, RNAi screens have limitations in the number of genes they are able to simultaneously target. Beyond investigations in cancer treatment, this tool should prove useful in investigating the wide spectrum of diseases who's patterns of cellular dysregulation play a role in determining patterns of drug sensitivity.

Bibliography

Adzhubei, Ivan A., Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. 2010. "A Method and Server for Predicting Damaging Missense Mutations." *Nature Methods* 7 (4) (April 1): 248–249. doi:10.1038/nmeth0410-248.

Bauer, S, P N Robinson, and J Gagneur. 2011. "Model-based Gene Set Analysis for Bioconductor." *Bioinformatics* 27 (13) (June): 1882–1883. doi:10.1093/bioinformatics/btr296.

Brakebusch, C, and R Fässler. 2005. "Beta 1 Integrin Function in Vivo: Adhesion, Migration and More." *Cancer Metastasis Reviews* 24 (3) (September): 403–411. doi:10.1007/s10555-005-5132-5.

Cheung, H W, G S Cowley, B A Weir, J S Boehm, S Rusin, J A Scott, A East, L D Ali, P H Lizotte, and T C Wong. 2011. "Systematic Investigation of Genetic Vulnerabilities Across Cancer Cell Lines Reveals Lineage-specific Dependencies in Ovarian Cancer." *Proceedings of the National Academy of Sciences* 108 (30): 12372–12377. doi:10.1073/pnas.1109363108/-/DCSupplemental/pnas.201109363SI.pdf.

Cline, Melissa S, Michael Smoot, Ethan Cerami, Allan Kuchinsky, Nerius Landys, Chris Workman, Rowan Christmas, et al. 2007. "Integration of Biological Networks and Gene Expression Data Using Cytoscape." *Nature Protocols* 2 (10) (October): 2366–2382. doi:10.1038/nprot.2007.324.

Demir, Emek, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, et al. 2010. "The BioPAX Community Standard for Pathway Data Sharing." *Nature Biotechnology* 28 (9) (September): 935–942. doi:10.1038/nbt.1666.

Druker, B J, S Tamura, E Buchdunger, S Ohno, G M Segal, S Fanning, J Zimmermann, and N B Lydon. 1996. "Effects of a Selective Inhibitor of the Abl Tyrosine Kinase on the Growth of Bcr-Abl Positive Cells." *Nature Medicine* 2 (5): 561–566.

Druker, Brian J., Moshe Talpaz, Debra J. Resta, Bin Peng, Elisabeth Buchdunger, John M. Ford, Nicholas B. Lydon, et al. 2001. "Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia." *New England Journal of Medicine* 344 (14): 1031–1037. doi:10.1056/NEJM200104053441401.

Early Breast Cancer Trialists' Collaborative Group. 1998. "Tamoxifen for Early Breast Cancer: An Overview of the Randomised Trials." *The Lancet* 351 (9114) (May 16): 1451–1467. doi:10.1016/S0140-6736(97)11423-4.

Echeverri, Christophe J, and Norbert Perrimon. 2006. "High-throughput RNAi Screening in Cultured Cells: a User's Guide." *Nature Reviews Genetics* 7 (5) (April): 373–384. doi:10.1038/nrg1836.

Eifert, Cheryl, and R Scott Powers. 2012. "From Cancer Genomes to Oncogenic Drivers, Tumour Dependencies and Therapeutic Targets." *Nature Publishing Group* 12 (8) (June): 572–578. doi:10.1038/nrc3299.

Gentleman, Robert C, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, et al. 2004. "Bioconductor: Open Software

Development for Computational Biology and Bioinformatics." *Genome Biology* 5 (10): R80. doi:10.1186/gb-2004-5-10-r80.
Goodsell, D S. 1999. "The Molecular Perspective: The Ras Oncogene." *The Oncologist* 4 (3): 263–264.
Grossmann, V., S. Schnittger, F. Poetzinger, A. Kohlmann, A. Stiel, C. Eder, A. Fasan, W. Kern, T. Haferlach, and C. Haferlach. 2013. "High Incidence of RAS Signalling Pathway Mutations in MLL-rearranged Acute Myeloid." *Leukemia* 27 (9) (March 28): 1933–1936. doi:10.1038/leu.2013.90.
Hammerman, Peter S, Michael S Lawrence, Douglas Voet, Rui Jing, Kristian Cibulskis, Andrey Sivachenko, Petar Stojanov, et al. 2012. "Comprehensive Genomic Characterization of Squamous Cell Lung Cancers." *Nature* 489 (7417) (September): 519–525. doi:10.1038/nature11404.
Hynes, Nancy E, and Gwen MacDonald. 2009. "ErbB Receptors and Signaling Pathways in Cancer." *Current Opinion in Cell Biology* 21 (2) (April): 177–184. doi:10.1016/j.ceb.2008.12.010.
Iorns, Elizabeth, Christopher J Lord, Nicholas Turner, and Alan Ashworth. 2007. "Utilizing RNA Interference to Enhance Cancer Drug Discovery." *Nature Reviews Drug Discovery* 6 (7) (July): 556–568. doi:10.1038/nrd2355.
Jordan, V Craig. 2003. "Tamoxifen: a Most Unlikely Pioneering Medicine." *Nature Reviews. Drug Discovery* 2 (3) (March): 205–213. doi:10.1038/nrd1031.
Koboldt, Daniel C, Robert S Fulton, Michael D McLellan, Heather Schmidt, Joelle Kalicki-Veizer, Joshua F McMichael, Lucinda L Fulton, et al. 2012. "Comprehensive Molecular Portraits of Human Breast Tumours." *Nature* 490 (7418) (September): 61–70. doi:10.1038/nature11412.
Kulesz-Martin, M F, J Lagowski, Susan Olson, Aaron Wortham, Toni West, George Thomas, Christopher Ryan, and Jeffrey W Tyner. 2013. "A Molecular Case Report: Functional Assay of Tyrosine Kinase Inhibitors in Cells from a Patient's Primary Renal Cell Carcinoma." *Cancer Biology & Therapy* 14 (2). http://migrate.landesbioscience.com/journals/cbt/article/22960/?show_full_text=true&.
Loriaux, M M, R L Levine, J W Tyner, S Frohling, C Scholl, E P Stoffregen, G Wernig, et al. 2008. "High-throughput Sequence Analysis of the Tyrosine Kinome in Acute Myeloid Leukemia." *Blood* 111 (9) (May): 4788–4796. doi:10.1182/blood-2007-07-101394.
Martín-Subero, José Ignacio, Lana Harder, Stefan Gesk, Robert Schoch, Francisco Javier Novo, Werner Grote, María José Calasanz, Brigitte Schlegelberger, and Reiner Siebert. 2001. "Amplification of ERBB2, RARA, and TOP2A Genes in a Myelodysplastic Syndrome Transforming to Acute Myeloid Leukemia." *Cancer Genetics and Cytogenetics* 127 (2) (June): 174–176. doi:10.1016/S0165-4608(00)00431-3.
McLendon, Roger, Allan Friedman, Darrell Bigner, Erwin G. Van Meir, Daniel J. Brat, Gena M. Mastrogianakis, Jeffrey J. Olson, et al. 2008. "Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways." *Nature* 455 (7216) (October 23): 1061–1068. doi:10.1038/nature07385.
"Mutation Annotation Format (MAF) Specification - TCGA - National Cancer Institute - Confluence Wiki." 2013. Accessed June 12.

https://wiki.nci.nih.gov/display/TCGA/Mutation+Annotation+Format+(MAF)+Specification.

Network, The Cancer Genome Atlas. 2012. "Comprehensive Molecular Characterization of Human Colon and Rectal Cancer." *Nature* 487 (7407) (July): 330–337. doi:10.1038/nature11252.

Ng, S, E A Collisson, A Sokolov, T Goldstein, A Gonzalez-Perez, N Lopez-Bigas, C Benz, D Haussler, and J M Stuart. 2012. "PARADIGM-SHIFT Predicts the Function of Mutations in Multiple Cancers Using Pathway Impact Analysis." *Bioinformatics* 28 (18) (September): i640–i646. doi:10.1093/bioinformatics/bts402.

Sales, Gabriele, Enrica Calura, Duccio Cavalieri, and Chiara Romualdi. 2012. "Graphite - a Bioconductor Package to Convert Pathway Topology to Gene Network." *BMC Bioinformatics* 13 (1) (January): 20. doi:10.1186/1471-2105-13-20.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13 (11) (November 1): 2498–2504. doi:10.1101/gr.1239303.

Sherry, Stephen T, Minghong Ward, and Karl Sirotkin. 1999. "dbSNP—database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation." *Genome Research* 9 (8): 677–679. doi:10.1101/gr.9.8.677.

Tyner, J W, W F Yang, A Bankhead, G Fan, L B Fletcher, J Bryant, J M Glover, et al. 2012. "Kinase Pathway Dependence in Primary Human Leukemias Determined by Rapid Inhibitor Screening." *Cancer Research* (October). doi:10.1158/0008-5472.CAN-12-1906. http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-12-1906.

Tyner, J. W., M. M. Loriaux, H. Erickson, C. A. Eide, J. Deininger, M. MacPartlin, S. G. Willis, et al. 2008. "High-throughput Mutational Screen of the Tyrosine Kinome in Chronic Myelomonocytic." *Leukemia* 23 (2) (July 10): 406–409. doi:10.1038/leu.2008.187.

Vaske, C J, S C Benz, J Z Sanborn, D Earl, C Szeto, J Zhu, D Haussler, and J M Stuart. 2010. "Inference of Patient-specific Pathway Activities from Multi-dimensional Cancer Genomics Data Using PARADIGM." *Bioinformatics* 26 (12) (June): i237–i245. doi:10.1093/bioinformatics/btq182.

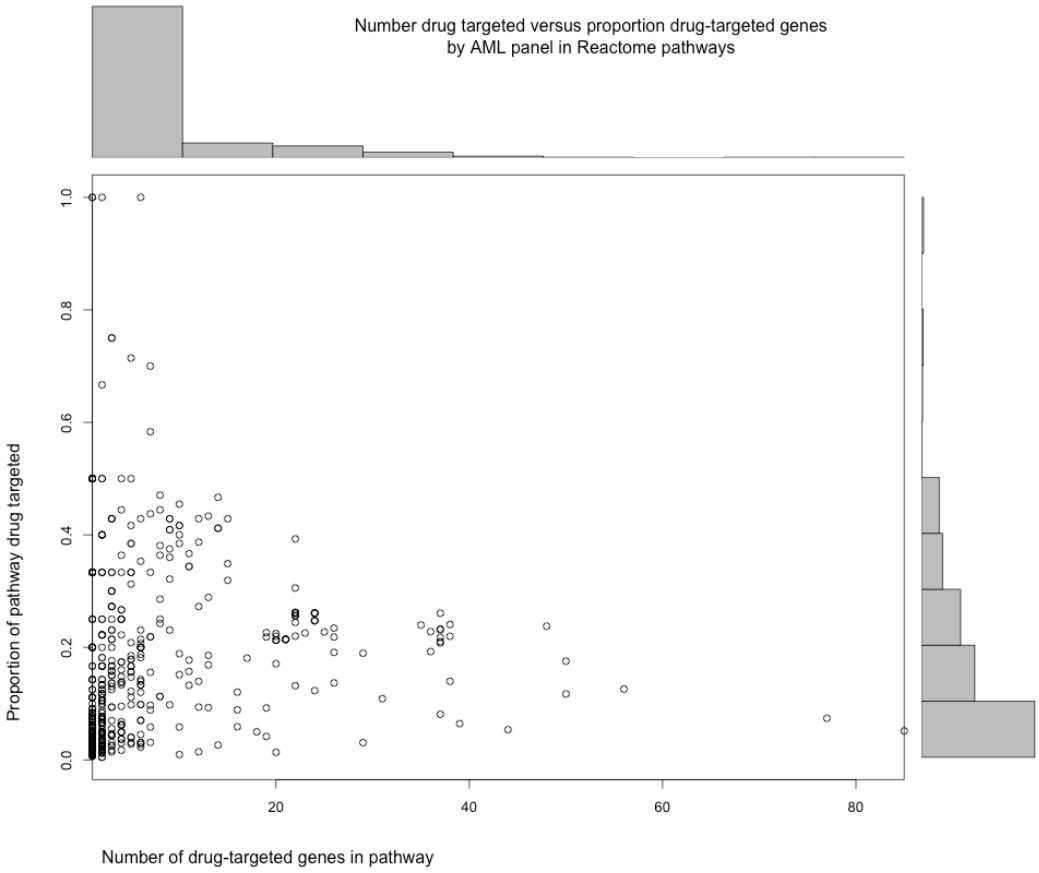
Verhaak, Roel G W, Pablo Tamayo, Ji-Yeon Yang, Diana Hubbard, Hailei Zhang, Chad J Creighton, Sian Fereday, et al. 2013. "Prognostically Relevant Gene Signatures of High-grade Serous Ovarian Carcinoma." *The Journal of Clinical Investigation* 123 (1) (January 2): 517–525. doi:10.1172/JCI65833.

Wishart, David S, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. 2006. "DrugBank: a Comprehensive Resource for in Silico Drug Discovery and Exploration." *Nucleic Acids Research* 34 (Database issue) (January 1): D668–672. doi:10.1093/nar/gkj067.

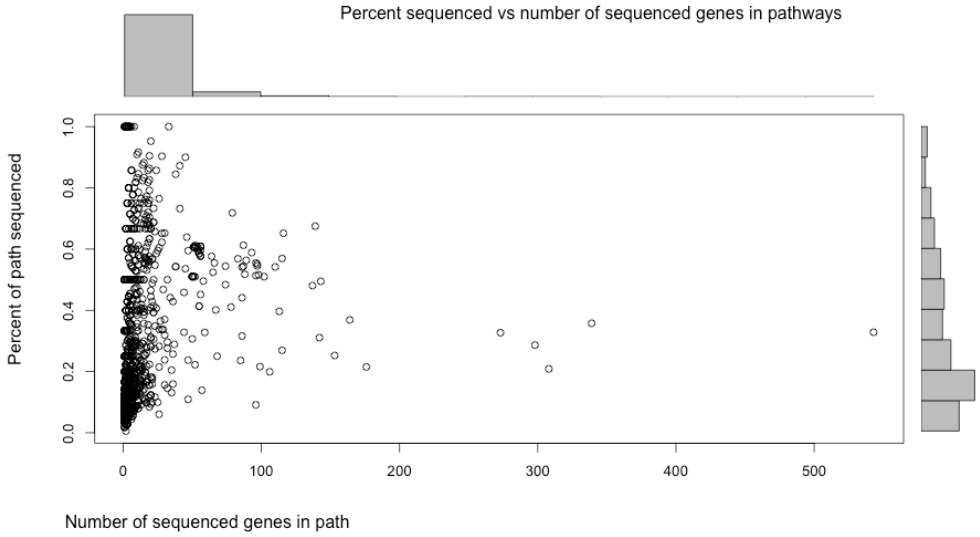
Yonesaka, Kimio, Kreshnik Zejnullahu, Isamu Okamoto, Taroh Satoh, Federico Cappuzzo, John Souglakos, Dalia Ercan, et al. 2011. "Activation of ERBB2 Signaling Causes Resistance to the EGFR-directed Therapeutic Antibody

Cetuximab." *Science Translational Medicine* 3 (99) (September 7): 99ra86. doi:10.1126/scitranslmed.3002442.

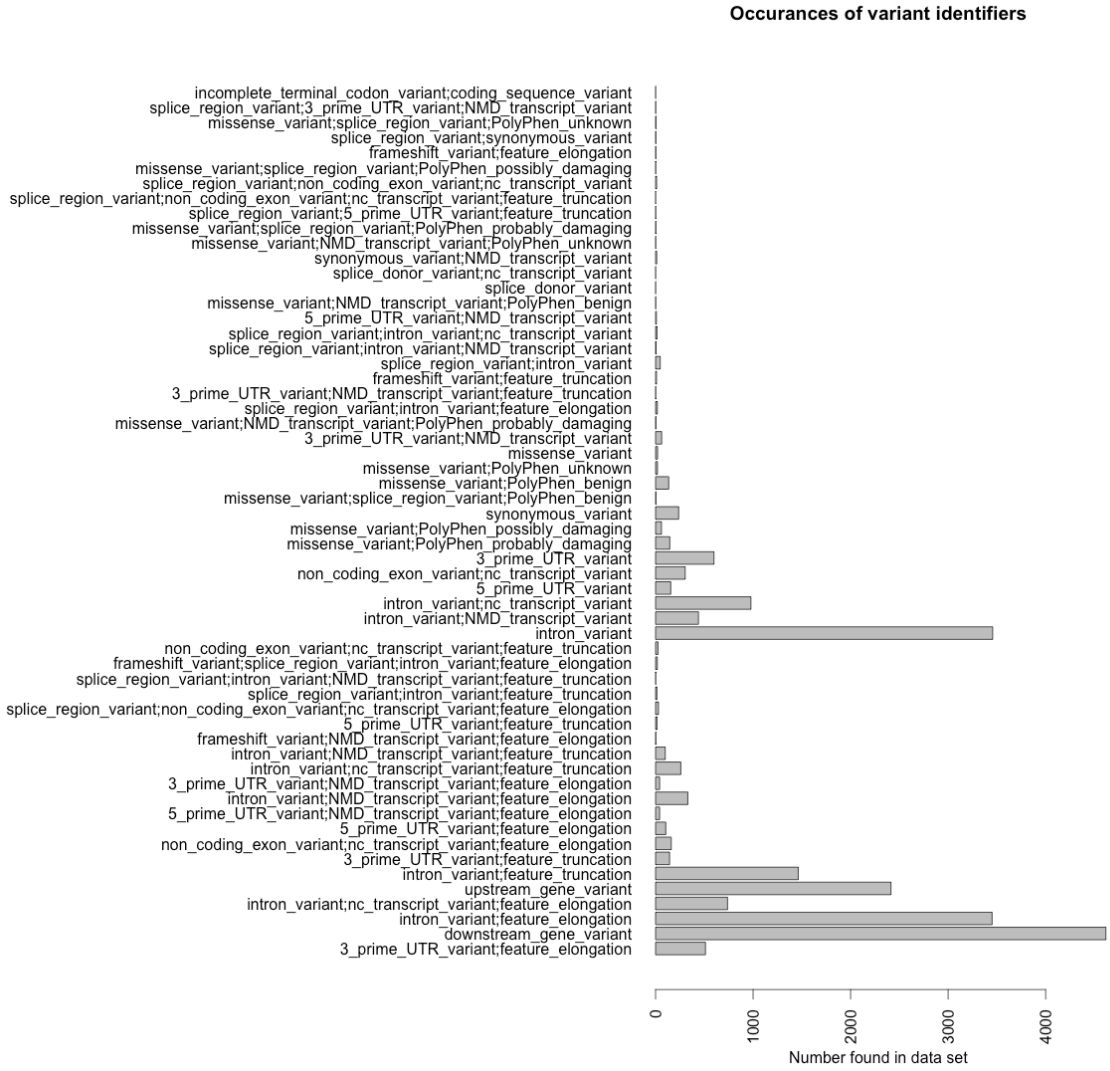
Appendix



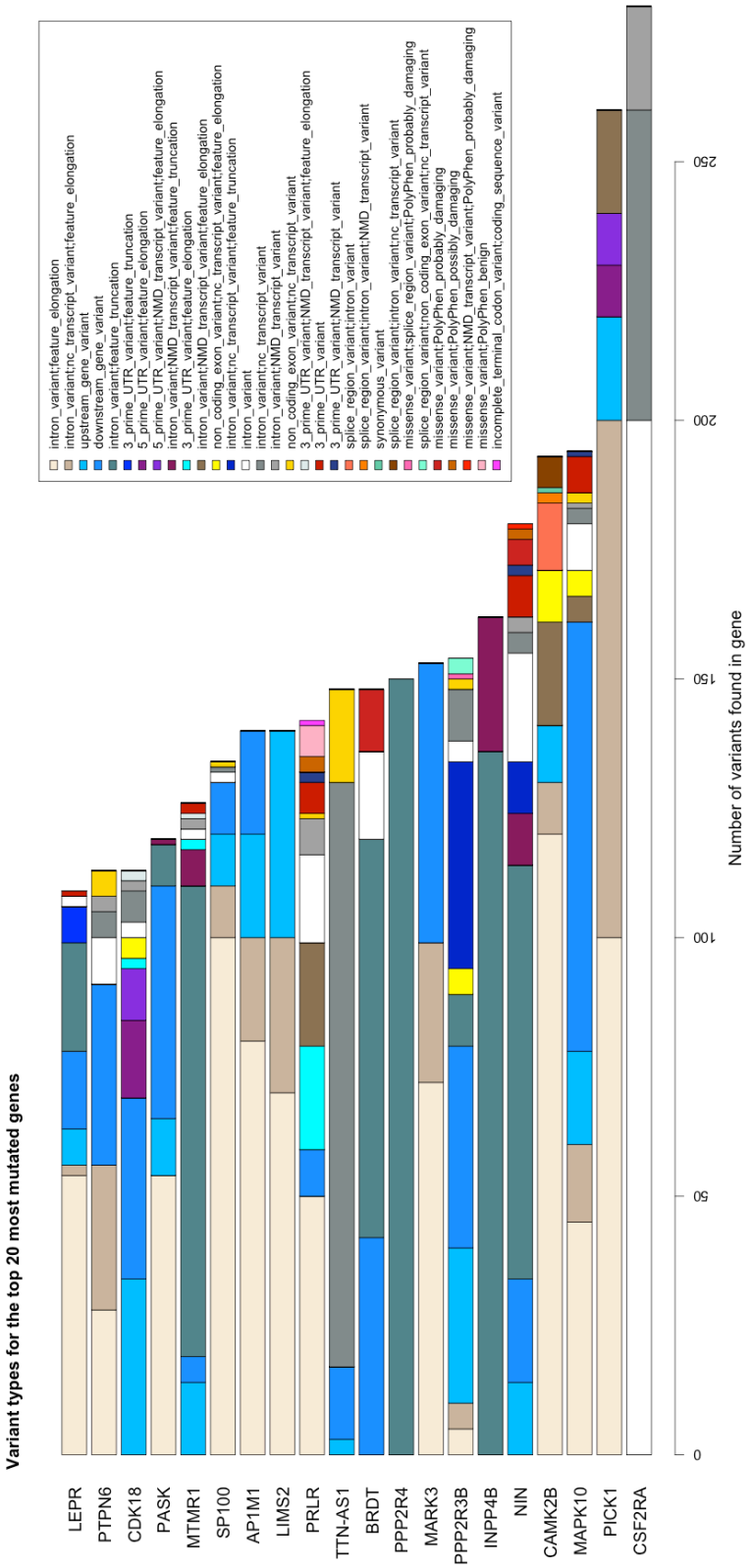
Supplementary figure 1: Comparison of the distributions of the numbers genes to the proportion of genes in Reactome pathways targeted by the AML panel from Tyner et al 2012.



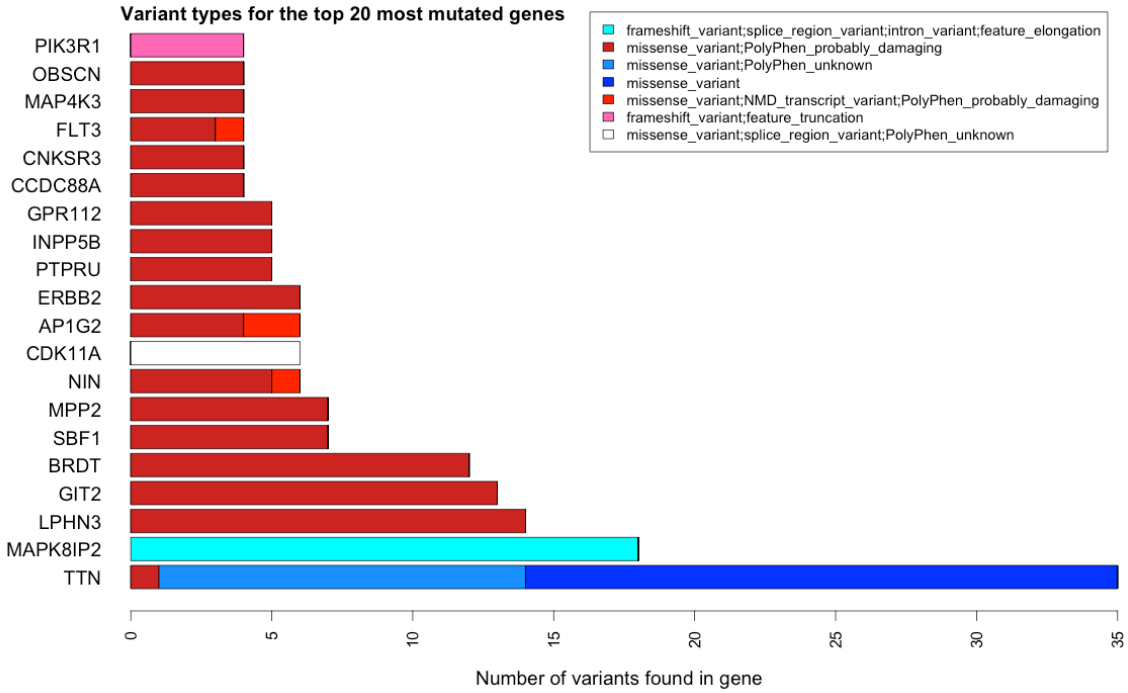
Supplementary figure 2: Percent sequenced, versus number of sequenced genes in pathway for the sequence capture gene variant data.



Supplementary figure 3: Distribution of variant types found in the sequence capture data for AML



Supplementary figure 4: Distribution of variant types in the genes found to have the top 20 highest numbers of variants in the AML sequence capture data.



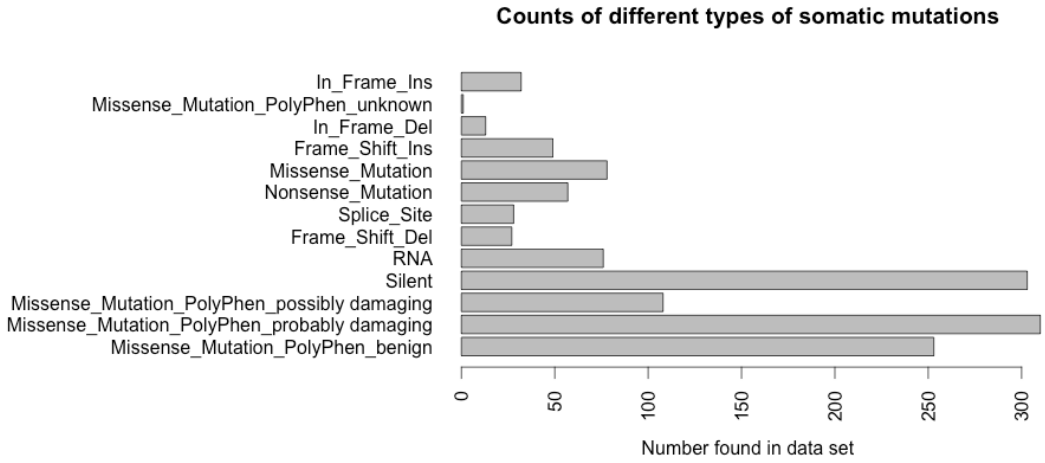
Supplementary figure 5: Distribution of the variant types in the top 20 most aberrational genes found in the AML sequence capture data, after all filtering steps.

Supplementary table 1: Variant types selected to indicate valid aberrations and variant types filtered out. Individual variant types are composed of a set of variant terms, which are separated by semicolons, while types are separated by semicolons and spaces.

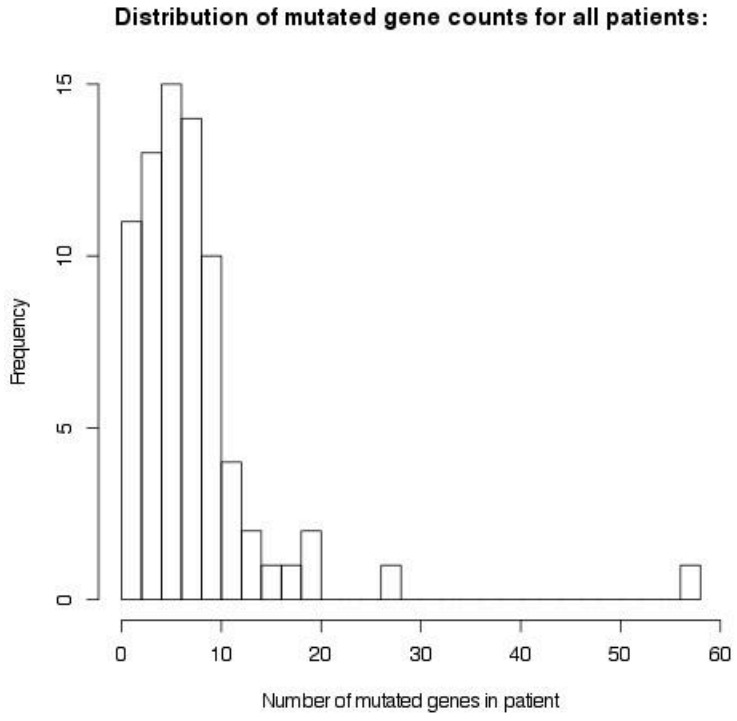
Variant types selected as indicating genes to be in an aberrational state	missense_variant;PolyPhen_probably_damaging; splice_acceptor_variant; splice_acceptor_variant;nc_transcript_variant; missense_variant;NMD_transcript_variant;PolyPhen_probably_damaging; missense_variant;PolyPhen_unknown; missense_variant;splice_region_variant;PolyPhen_probably_damaging; frameshift_variant;splice_region_variant;intron_variant;feature_elongation; splice_donor_variant; splice_donor_variant;nc_transcript_variant; stop_gained; stop_gained;NMD_transcript_variant; frameshift_variant;feature_elongation; missense_variant;NMD_transcript_variant;PolyPhen_unknown; missense_variant; frameshift_variant;feature_truncation; frameshift_variant;NMD_transcript_variant;feature_truncation; frameshift_variant;splice_region_variant;feature_truncation; missense_variant;splice_region_variant;PolyPhen_unknown; incomplete_terminal_codon_variant;coding_sequence_variant; splice_acceptor_variant;NMD_transcript_variant; splice_donor_variant;feature_elongation; splice_donor_variant;nc_transcript_variant;feature_elongation; inframe_insertion; splice_acceptor_variant;nc_transcript_variant;feature_elongation; missense_variant;splice_region_variant;NMD_transcript_variant;PolyPhen_unknown; splice_acceptor_variant;feature_truncation; splice_acceptor_variant;nc_transcript_variant;feature_truncation; stop_lost; stop_gained;splice_region_variant
Variant types selected as not qualifying valid aberrations	3_prime_UTR_variant;feature_elongation; downstream_gene_variant; intron_variant;feature_elongation; intron_variant;nc_transcript_variant;feature_elongation; upstream_gene_variant; intron_variant;feature_truncation; 3_prime_UTR_variant;feature_truncation; non_coding_exon_variant;nc_transcript_variant;feature_elongation; 5_prime_UTR_variant;feature_elongation; 5_prime_UTR_variant;NMD_transcript_variant;feature_elongation; intron_variant;NMD_transcript_variant;feature_elongation; 3_prime_UTR_variant;NMD_transcript_variant;feature_elongation; intron_variant;nc_transcript_variant;feature_truncation; intron_variant;NMD_transcript_variant;feature_truncation; frameshift_variant;NMD_transcript_variant;feature_elongation; 5_prime_UTR_variant;feature_truncation; splice_region_variant;non_coding_exon_variant;nc_transcript_variant;feature_elongation; splice_region_variant;intron_variant;feature_truncation; splice_region_variant;intron_variant;NMD_transcript_variant;feature_truncation; non_coding_exon_variant;nc_transcript_variant;feature_truncation; intron_variant; intron_variant;NMD_transcript_variant; intron_variant;nc_transcript_variant; 5_prime_UTR_variant; non_coding_exon_variant;nc_transcript_variant; 3_prime_UTR_variant; missense_variant;PolyPhen_possibly_damaging; synonymous_variant; missense_variant;splice_region_variant;PolyPhen_benign; missense_variant;PolyPhen_benign; 3_prime_UTR_variant;NMD_transcript_variant; splice_region_variant;intron_variant;feature_elongation; 3_prime_UTR_variant;NMD_transcript_variant;feature_truncation; splice_region_variant;intron_variant; splice_region_variant;intron_variant;NMD_transcript_variant; splice_region_variant;intron_variant;nc_transcript_variant; 5_prime_UTR_variant;NMD_transcript_variant; missense_variant;NMD_transcript_variant;PolyPhen_benign; synonymous_variant;NMD_transcript_variant; splice_region_variant;5_prime_UTR_variant;feature_truncation; splice_region_variant;non_coding_exon_variant;nc_transcript_variant;feature_truncation; splice_region_variant;non_coding_exon_variant;nc_transcript_variant; missense_variant;splice_region_variant;PolyPhen_possibly_damaging; splice_region_variant;synonymous_variant; splice_region_variant;3_prime_UTR_variant;NMD_transcript_variant

Supplementary table 2: Variant types retained as valid aberrations (top) and discarded (bottom) from the AML somatic mutation data analyzed in use case 2.

Retained these variant types as valid aberrations	Missense_Mutation_PolyPhen_probably damaging; Frame_Shift_Del; Splice_Site; Nonsense_Mutation; Missense_Mutation; Frame_Shift_Ins; In_Frame_Del; Missense_Mutation_PolyPhen_unknown; In_Frame_Ins
Filtered out these variant types	Missense_Mutation_PolyPhen_benign; Missense_Mutation_PolyPhen_possibly damaging; Silent; RNA



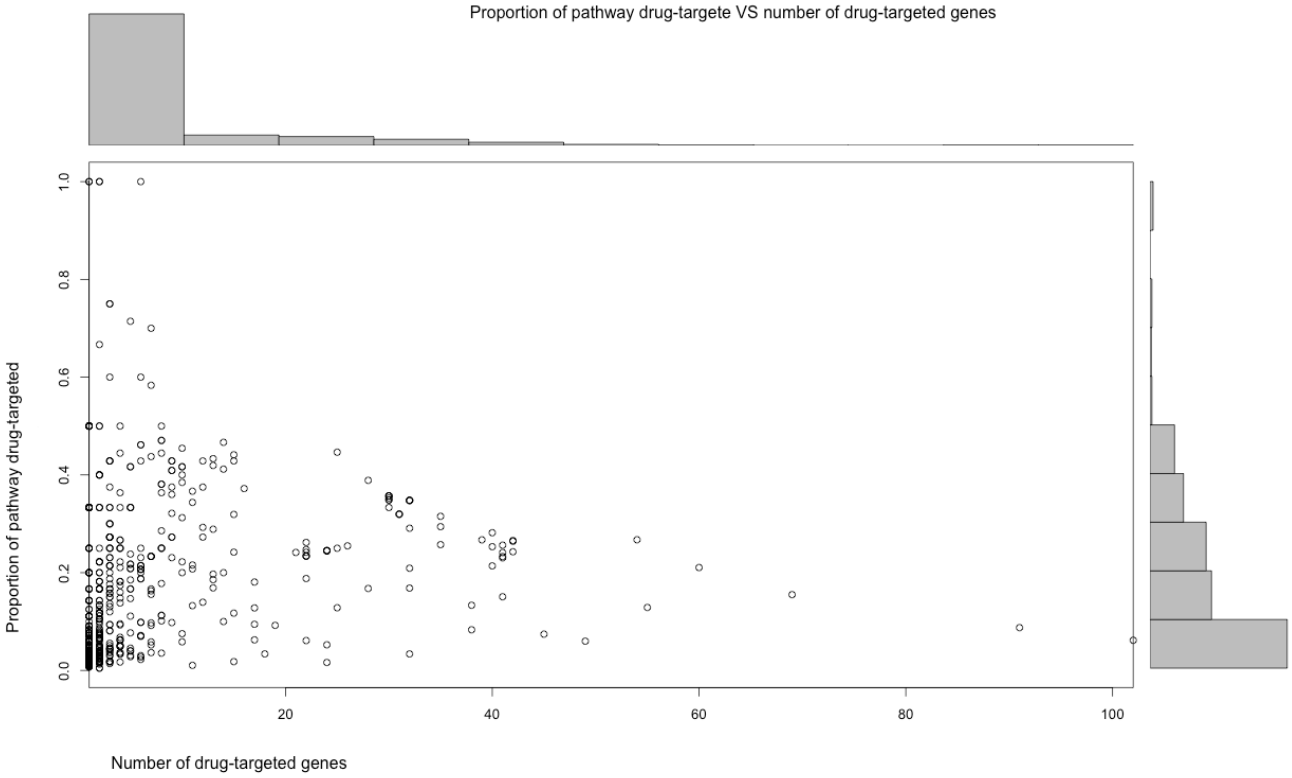
Supplementary figure 6: Distributions of all variant types found in the somatic mutation data



Supplementary figure 7: Distribution of aberrations per patient found in the AML somatic mutation data, after all filtering.

Supplementary table 3: Variant types selected to imply valid aberration in the examination of HNSCC data from TCGA

Retained these aberration types as valid aberrations	Missense_Mutation_PolyPhen_probably damaging; Splice_Site; Nonsense_Mutation; Frame_Shift_Del; Missense_Mutation; In_Frame_Del; Frame_Shift_Ins; In_Frame_Ins; Nonstop_Mutation; Missense_Mutation_PolyPhen_unknown
Filtered out these variatn types	Missense_Mutation_PolyPhen_possibly damaging; Silent; Missense_Mutation_PolyPhen_benign; RNA; 5'Flank; Translation_Start_Site; IGR



Supplementary figure 6: Comparison of number of drug targeted genes to the proportion of the path targeted by the current panel design for HNSCC

Supplementary materials

Column header abreviations

T P L drg Scn	testable path length for the drug screen analysis
# D S Gen	Number of drug sensitive genes
P P D S	proportion of pathway genes which were found to be drug sensitive
P. co. w D sens in pth	proportion of cohort w drug sensitive gene in path
# pat W D S G in Pth	Num patient with drug sensitive gene .s. in path
C D S x in Pths x coh	count of drug sensitive genes in path cross cohort
F P L	full path length
T P L Ab	testable path length for aberration analysis
ab. gen.	aberrational genes
p. ab.	percent of pathway found to be aberrational
P. co. w ab in pth	proportion of cohort w aberrational gene in path
# pat w ab in pth	Number of patients with aberrational gene(s) in path
# ab x co	Number of aberrational genes in path cross cohort
Mx O G	number of patients gene is mutated in for the gene that is most frequently mutated in the cohort
Mi O G	number of patients gene is mutated in for the gene that is least frequently mutated in the cohort
Hyp G	Hyper geometric p-value
Hyp G FDR	Hyper geometric p-value with False discovery rate correction

AAK1
ACVR1
ACVRL1
ADCK3
ADCK4
ALK
ANKK1
NUAK1
AURKC
AXL
BMP2K
BRSK1
BRSK2
CAMK1
CAMK1D
CAMK1G
CAMKK1
CAMKK2
CDK3
CIT
CLK1
CLK2
CLK3
CLK4
CSF1R
CSNK1A1L
CSNK1G1
CSNK1G3
DCLK1
DCLK2
DCLK3
DDR1
DDR2
MAP3K12
DMPK
CDC42BPG
STK17A
STK17B
DYRK1B
EPHA1
EPHA2
EPHA3
EPHA4
EPHA5
EPHA6
EPHA7
EPHA8
EPHB1
EPHB3
EPHB4
MAPK6
MAPK4
MAPK15
FLT3
FRK
EIF2AK4
INSRR
STK10
LTK
MAP3K4
MAP3K5
MAP4K1
MAP4K3
MAP4K4
MAP4K5
MAPKAPK5
MARK1
MARK2
MARK3
MARK4
MELK
MKNK2
MYLK3
MAP3K9
MAP3K10
MAP3K11
CDC42BPA
CDC42BPB
MST1
MUSK
MYLK2
MYO3A
MYO3B
STK38L
NEK1
NEK5
NEK6
NEK7
NEK9
NLK
CDK16
CDK17
CDK18
CDK14
PIM1
PIM2
PIM3
PKN1
PKN2
PLK3
PRKD2
PRKD3
PRKX
PTK6
RET
RIOK1
RIOK3
ROS1
MYLK4
SLK
NUAK2
SIK1
SIK2
SRMS
SRPK1
SRPK2
STK16
STK33
STK36
TIE1
TLK1
TLK2
TNIK
TNK1
TNK2
TNNI3K
NTRK3
TSSK1B
TTK
TXK
TYRO3
STK32B
STK32C
STK25
ZAK

Paths enriched in aberration. Showing top 86 pathways, sorted by number of aberrational genes

path id	F P L	T P L ab	variant genes	proportion variant	proportion of cohort w variant gene in path	Num patient with variant gene (s) in path	count of variant genes in path cross cohort	Mx O G	Mi O G	Hyp G	Hyp G FDR	mgsa probability estimate	mgsa std error
Signaling Pathways	1874	543	14	0.026	0.7	7	14	1	0	0.053	0.101	0	0
GPCR downstream signaling	909	273	8	0.029	0.6	6	8	1	0	0.063	0.11	0	0
Signaling by GPCR	1022	339	8	0.024	0.6	6	8	1	0	0.19	0.253	0	0
Signaling by NGF	374	137	5	0.036	0.3	3	5	1	0	0.042	0.09	0	0
Immune System	1335	298	5	0.017	0.3	3	5	1	0	0.544	0.547	0	0
Metabolism	1569	308	5	0.016	0.4	4	5	1	0	0.581	0.581	0	0
G alpha (q) signalling events	181	116	4	0.034	0.4	4	4	1	0	0.068	0.113	0	0
Gastrin-CREB signalling pathway via PKC and MAPK	210	139	4	0.029	0.4	4	4	1	0	0.125	0.179	0	0
Adaptive Immune System	856	153	4	0.026	0.3	3	4	1	0	0.168	0.231	0	0
Disease	1105	176	4	0.023	0.4	4	4	1	0	0.252	0.316	0	0
Rho GTPase cycle	212	19	3	0.158	0.1	1	3	1	0	0	0.015	0.189	0.013
Signaling by Rho GTPases	212	19	3	0.158	0.1	1	3	1	0	0	0.015	0.193	0.012
p75 NTR receptor-mediated signalling	143	23	3	0.13	0.2	2	3	1	0	0.001	0.02	0.049	0.003
TCR signaling	77	29	3	0.103	0.3	3	3	1	0	0.002	0.02	0.034	0.003
G alpha (12/13) signalling events	118	34	3	0.088	0.2	2	3	1	0	0.003	0.026	0.006	0
G alpha (s) signalling events	126	65	3	0.046	0.2	2	3	1	0	0.034	0.078	0	0
Downstream Signaling Events Of B Cell Receptor (BCR)	179	67	3	0.045	0.2	2	3	1	0	0.037	0.082	0	0

Signaling by ERBB4	175	74	3	0.041	0.2	2	3	1	0	0.051	0.1	0	0
B Cell Activation	344	78	3	0.038	0.2	2	3	1	0	0.06	0.107	0	0
Cell Cycle, Mitotic	371	85	3	0.035	0.4	4	4	2	0	0.078	0.121	0	0
Signaling by ERBB2	171	86	3	0.035	0.3	3	3	1	0	0.081	0.124	0	0
Signaling by SCF-KIT	155	87	3	0.034	0.3	3	3	1	0	0.084	0.128	0	0
Gene Expression	1092	96	3	0.031	0.4	4	4	2	0	0.111	0.164	0	0
Cell Cycle	457	99	3	0.03	0.4	4	4	2	0	0.121	0.174	0	0
Cytokine Signaling in Immune system	349	113	3	0.027	0.2	2	3	1	0	0.172	0.234	0	0
Developmental Biology	450	115	3	0.026	0.2	2	3	1	0	0.18	0.243	0	0
GPCR ligand binding	292	143	3	0.021	0.2	2	3	1	0	0.302	0.32	0	0
Innate Immune System	649	164	3	0.018	0.2	2	3	1	0	0.401	0.408	0	0
Nonsense-Mediated Decay	107	11	2	0.182	0.3	3	3	2	0	0.001	0.02	0.119	0.011
Nonsense Mediated Decay Enhanced by the Exon Junction Complex	107	11	2	0.182	0.3	3	3	2	0	0.001	0.02	0.119	0.011
NRAGE signals death through JNK	90	13	2	0.154	0.1	1	2	1	0	0.002	0.02	0.033	0.002
Cell death signalling via NRAGE, NRIF and NADE	119	16	2	0.125	0.1	1	2	1	0	0.003	0.026	0.018	0.001
Downstream TCR signaling	60	20	2	0.1	0.2	2	2	1	0	0.006	0.035	0.016	0.002
Cyclin D associated events in G1	39	21	2	0.095	0.3	3	3	2	0	0.007	0.035	0.014	0.002
G1 Phase	39	21	2	0.095	0.3	3	3	2	0	0.007	0.035	0.013	0.001
Semaphorin interactions	66	23	2	0.087	0.2	2	2	1	0	0.009	0.038	0.009	0.001

CD28 co-stimulation	33	28	2	0.071	0.2	2	2	1	0	0.016	0.047	0.003	0
Metabolism of mRNA	221	32	2	0.062	0.3	3	3	2	0	0.022	0.058	0.002	0
Synthesis of PIPs at the plasma membrane	33	33	2	0.061	0.2	2	2	1	0	0.024	0.061	0.001	0
Metabolism of RNA	269	35	2	0.057	0.3	3	3	2	0	0.028	0.069	0.001	0
Costimulation by the CD28 family	93	38	2	0.053	0.2	2	2	1	0	0.035	0.08	0	0
Mitotic G1-G1/S phases	139	44	2	0.045	0.3	3	3	2	0	0.051	0.1	0	0
Constitutive PI3K/AKT Signaling in Cancer	94	45	2	0.044	0.2	2	2	1	0	0.054	0.101	0	0
PI Metabolism	50	45	2	0.044	0.2	2	2	1	0	0.054	0.101	0	0
HIV Infection	424	47	2	0.043	0.2	2	2	1	0	0.06	0.107	0	0
PI3K/AKT Signaling in Cancer	109	50	2	0.04	0.2	2	2	1	0	0.07	0.113	0	0
PI-3K cascade	109	50	2	0.04	0.2	2	2	1	0	0.07	0.113	0	0
PI3K events in ERBB2 signaling	109	50	2	0.04	0.2	2	2	1	0	0.07	0.113	0	0
PI3K events in ERBB4 signaling	109	50	2	0.04	0.2	2	2	1	0	0.07	0.113	0	0
PIP3 activates AKT signaling	109	50	2	0.04	0.2	2	2	1	0	0.07	0.113	0	0
PI3K/AKT activation	111	51	2	0.039	0.2	2	2	1	0	0.074	0.118	0	0
GAB1 signalosome	113	52	2	0.038	0.2	2	2	1	0	0.077	0.121	0	0
Phospholipid metabolism	132	55	2	0.036	0.2	2	2	1	0	0.088	0.134	0	0
Signaling by Interleukins	112	79	2	0.025	0.2	2	2	1	0	0.196	0.26	0	0
Downstream signaling of activated FGFR	157	83	2	0.024	0.2	2	2	1	0	0.217	0.284	0	0
Axon guidance	274	86	2	0.023	0.2	2	2	1	0	0.233	0.299	0	0

Platelet activation, signaling and aggregation	198	86	2	0.023	0.7	7	7	6	0	0.233	0.299	0	0
DAP12 signaling	169	87	2	0.023	0.2	2	2	1	0	0.238	0.304	0	0
DAP12 interactions	185	88	2	0.023	0.2	2	2	1	0	0.243	0.309	0	0
Signaling by FGFR	171	89	2	0.022	0.2	2	2	1	0	0.249	0.314	0	0
Downstream signal transduction	170	93	2	0.022	0.2	2	2	1	0	0.27	0.319	0	0
Signaling by PDGF	199	96	2	0.021	0.2	2	2	1	0	0.287	0.319	0	0
Signaling by FGFR in disease	190	96	2	0.021	0.2	2	2	1	0	0.287	0.319	0	0
Signaling by EGFR in Cancer	193	97	2	0.021	0.2	2	2	1	0	0.292	0.319	0	0
Signaling by EGFR	191	97	2	0.021	0.2	2	2	1	0	0.292	0.319	0	0
Peptide ligand-binding receptors	191	98	2	0.02	0.2	2	2	1	0	0.298	0.319	0	0
Class A/1 (Rhodopsin-like receptors)	202	102	2	0.02	0.2	2	2	1	0	0.32	0.335	0	0
Metabolism of lipids and lipoproteins	554	106	2	0.019	0.2	2	2	1	0	0.342	0.356	0	0
NGF signalling via TRKA from the plasma membrane	225	115	2	0.017	0.2	2	2	1	0	0.393	0.402	0	0
Hemostasis	454	142	2	0.014	0.7	7	7	6	0	0.537	0.542	0	0
Regulation of PAK-2p34 activity by PS-GAP/RHG10	2	1	1	1	0.1	1	1	1	1	0	0	0.07	0.003
Stimulation of the cell death response by PAK-2p34	2	1	1	1	0.1	1	1	1	1	0	0	0.068	0.004

GRB7 events in ERBB2 signaling	6	2	1	0.5	0.1	1	1	1	0	0	0.015	0.061	0.004
Abortive elongation of HIV-1 transcript in the absence of Tat	24	3	1	0.333	0.1	1	1	1	0	0.001	0.02	0.035	0.002
Interleukin-1 processing	7	3	1	0.333	0.1	1	1	1	0	0.001	0.02	0.035	0.003
Elongation arrest and recovery	32	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.027	0.003
HIV-1 elongation arrest and recovery	32	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.025	0.002
Pausing and recovery of HIV-1 elongation	32	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.026	0.002
Pausing and recovery of Tat-mediated HIV-1 elongation	56	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.028	0.003
Tat-mediated HIV-1 elongation arrest and recovery	56	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.026	0.003
Regulation of Glucokinase by Glucokinase Regulatory Protein	31	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.042	0.005
Inflammasomes	23	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.027	0.002
Regulated proteolysis of p75NTR	18	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.028	0.002
GP1b-IX-V activation signalling	10	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.032	0.003
Nef and signal transduction	35	4	1	0.25	0.1	1	1	1	0	0.002	0.02	0.026	0.002
Striated Muscle Contraction	31	5	1	0.2	0.6	6	6	6	0	0.004	0.026	0.031	0.004

path id	F P L	T P L drg Scn	# D S Gen	P P D S	P. co. w D sens in pth	# pat W D S G in Pth	C D S x in Pths x coh	Mx O G	Mi O G	Hyp G	Hyp G FDR	F P L	T P L Ab	ab. gen	p. ab.	P. co. w ab in pth	# pat w ab in pth	# ab x co	Mx O G	Mi O G	Hyp G	Hyp G FDR
Activation of BIM and translocation to mitochondria	3	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Activation of BMF and translocation to mitochondria	3	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Phosphorylation of proteins involved in G1/S transition by active Cyclin E:Cdk2 complexes	4	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
c-src mediated regulation of Cx43 function and closure of gap junctions	4	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of gap junction activity	4	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G2 Phase	5	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GRB7 events in ERBB2 signaling	6	1	1	1	0.1	1	1	1	1	0	0	6	2	1	0.50	0.1	1	1	1	0	4E-04	0.015
Role of Abl in Robo-Slit signaling	9	2	2	1	0.6	6	7	5	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Interleukin-7 signaling	9	2	2	1	0.2	2	3	2	1	0	0	9	7	1	0.14	0.1	1	1	1	0	0.007	0.035

Downregulation of ERBB4 signaling	10	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PECAM1 interactions	12	5	5	1	0.3	3	8	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Platelet Adhesion to exposed collagen	13	2	2	1	0.3	3	4	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Interleukin-6 signaling	13	3	3	1	0.2	2	4	2	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of IFNG signaling	14	2	2	1	0.2	2	3	2	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Prolactin receptor signaling	15	1	1	1	0.2	2	2	2	2	0	0	15	8	1	0.13	0.1	1	1	1	0	0.01	0.038
Platelet sensitization by LDL	22	2	2	1	0.6	6	7	4	3	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Gap junction trafficking and regulation	22	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ADP signalling through P2Y purinoceptor 1	24	1	1	1	0.4	4	4	4	4	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G0 and Early G1	25	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of IFNA signaling	26	2	2	1	0.1	1	2	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NRIF signals cell death from the nucleus	27	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Signal amplification	30	1	1	1	0.4	4	4	4	4	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EGFR downregulation	31	1	1	1	0.2	2	2	2	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Activation of the pre-replicative complex	31	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Nef Mediated CD4 Down-regulation	36	1	1	1	0.2	2	2	2	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters	48	1	1	1	0.2	2	2	2	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Golgi Associated Vesicle Biogenesis	49	1	1	1	0.1	1	1	1	1	0	0	49	21	1	0.05	0.1	1	1	1	0	0.061	0.107
Nuclear signaling by ERBB4	50	2	2	1	0.2	2	3	2	1	0	0	50	11	1	0.09	0.1	1	1	1	0	0.018	0.051
CDK-mediated phosphorylation and removal of Cdc6	50	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Meiotic Recombination	54	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Clathrin derived vesicle budding	56	1	1	1	0.1	1	1	1	1	0	0	56	22	1	0.05	0.1	1	1	1	0	0.066	0.112
trans-Golgi Network Vesicle Budding	56	1	1	1	0.1	1	1	1	1	0	0	56	22	1	0.05	0.1	1	1	1	0	0.066	0.112
SCF(Skp2)-mediated degradation of p27/p21	56	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
p53-Dependent G1 DNA Damage Response	57	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

p53-Dependent G1/S DNA damage checkpoint	57	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Orc1 removal from chromatin	71	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Switching of origins to a post-replicative state	71	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Removal of licensing factors from origins	73	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of DNA replication	76	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DNA Replication Pre-Initiation	83	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M/G1 Transition	83	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Meiosis	85	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NRAGE signals death through JNK	90	1	1	1	0.1	1	1	1	1	0	0	90	13	2	0.15	0.1	1	2	1	0	0.002	0.02
Synthesis of DNA	96	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DNA Replication	104	1	1	1	0.1	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Interferon alpha/beta signaling	113	2	2	1	0.1	1	2	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cell death signalling via NRAGE, NRIF and NADE	119	1	1	1	0.1	1	1	1	1	0	0	119	16	2	0.13	0.1	1	2	1	0	0.003	0.026
Regulation of KIT signaling	16	7	6	0.857	0.5	5	13	5	0	7.07E-06	3.60E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Interleukin-3, 5 and GM-CSF signaling	46	13	7	0.538	0.6	6	12	3	0	0.000694	0.00345	46	38	1	0.03	0.1	1	1	1	0	0.168	0.231
Signaling by SCF-KIT	155	37	14	0.378	0.7	7	29	5	0	0.000709	0.00346	155	87	3	0.03	0.3	3	3	1	0	0.084	0.128
B Cell Activation	344	26	11	0.423	0.7	7	26	5	0	0.000782	0.00374	344	78	3	0.04	0.2	2	3	1	0	0.06	0.107
Antigen Activates B Cell Receptor Leading to Generation of Second Messengers	171	6	4	0.667	0.4	4	7	3	0	0.001249	0.0057	171	16	1	0.06	0.1	1	1	1	0	0.037	0.082
GRB2 events in ERBB2 signaling	33	11	6	0.545	0.3	3	8	2	0	0.001262	0.0057	33	20	1	0.05	0.1	1	1	1	0	0.056	0.101
SHC1 events in ERBB2 signaling	36	11	6	0.545	0.3	3	8	2	0	0.001262	0.0057	36	22	1	0.05	0.1	1	1	1	0	0.066	0.112
Nef and signal transduction	35	4	3	0.75	0.3	3	5	2	0	0.00136	0.00593	35	4	1	0.25	0.1	1	1	1	0	0.002	0.02
The role of Nef in HIV-1 replication and disease pathogenesis	55	4	3	0.75	0.3	3	5	2	0	0.00136	0.00593	55	18	1	0.06	0.1	1	1	1	0	0.046	0.093
Immune System	1335	77	22	0.286	1	10	45	5	0	0.001941	0.00831	1335	298	5	0.02	0.3	3	5	1	0	0.544	0.547
CTLA4 inhibitory signaling	28	9	5	0.556	0.3	3	8	3	0	0.00222	0.00918	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cell surface interactions at the vascular wall	93	9	5	0.556	0.3	3	8	3	0	0.00222	0.00918	93	26	1	0.04	0.1	1	1	1	0	0.089	0.134
Signaling by ERBB4	175	29	11	0.379	0.7	7	24	5	0	0.002734	0.01112	175	74	3	0.04	0.2	2	3	1	0	0.051	0.1
Signaling by ERBB2	171	37	13	0.351	0.7	7	26	5	0	0.002916	0.01166	171	86	3	0.03	0.3	3	3	1	0	0.081	0.124

path id	full path length	testable path length	aberrational genes	proportion aberrational	proportion of cohort w aberrational gene in path	Num patient with aberrational gene .s. in path	count of aberrational genes in path cross cohort	max in one gene	min in one gene	hyperg p value	hyperg p w FDR
Abortive elongation of HIV-1 transcript in the absence of Tat	24	3	1	0.33333333	0.1	1	1	1	0	0.00108	0.02022
Interleukin-1 processing	7	3	1	0.33333333	0.1	1	1	1	0	0.00108	0.02022
Regulation of Glucokinase by Glucokinase Regulatory Protein	31	4	1	0.25	0.1	1	1	1	0	0.00214	0.02022
Inflammasomes	23	4	1	0.25	0.1	1	1	1	0	0.00214	0.02022
Regulated proteolysis of p75NTR	18	4	1	0.25	0.1	1	1	1	0	0.00214	0.02022
Striated Muscle Contraction	31	5	1	0.2	0.6	6	6	6	0	0.00352	0.02617
Molecules associated with elastic fibres	24	5	1	0.2	0.1	1	1	1	0	0.00352	0.02617
Inhibition of replication initiation of damaged DNA by RB1/E2F1	13	5	1	0.2	0.1	1	1	1	0	0.00352	0.02617
TRAF6 mediated NF-kB activation	22	6	1	0.16666667	0.1	1	1	1	0	0.00522	0.03194

NF-kB is activated and signals survival	13	6	1	0.16666667	0.1	1	1	1	0	0.00522	0.03194
Orexin and neuropeptides FF and QRFP bind to their respective receptors	8	6	1	0.16666667	0.1	1	1	1	0	0.00522	0.03194
Relaxin receptors	8	6	1	0.16666667	0.1	1	1	1	0	0.00522	0.03194
Elastic fibre formation	35	7	1	0.14285714	0.1	1	1	1	0	0.00722	0.03493
Synthesis of PIPs at the late endosome membrane	10	8	1	0.125	0.1	1	1	1	0	0.00951	0.03806
Platelet degranulation	77	10	1	0.1	0.6	6	6	6	0	0.01493	0.04567
Hexose uptake	45	10	1	0.1	0.1	1	1	1	0	0.01493	0.04567
Glucose transport	43	10	1	0.1	0.1	1	1	1	0	0.01493	0.04567
Synthesis of IP2, IP, and Ins in the cytosol	11	10	1	0.1	0.1	1	1	1	0	0.01493	0.04567
Nonsense-Mediated Decay	107	11	2	0.18181818	0.3	3	3	2	0	0.00097	0.02022
Nonsense Mediated Decay Enhanced by the Exon Junction Complex	107	11	2	0.18181818	0.3	3	3	2	0	0.00097	0.02022
Rho GTPase cycle	212	19	3	0.15789474	0.1	1	3	1	0	0.00036	0.01519
Signaling by Rho GTPases	212	19	3	0.15789474	0.1	1	3	1	0	0.00036	0.01519

path id	full path length	testable path length	mutated genes	proportion mutated	proportion of cohort w mutated gene in path	Num patient with mutated gene (s) in path	count of mutated genes in path cross cohort	max in one gene	min in one gene	hyperg p value	hyperg p w FDR	mgsa probability estimate	mgsa std error
Cohesin Loading onto Chromatin	10	10	5	0.5	0.133333333	10	11	3	0	3.15E-08	1.64E-05	0.461084	0.01164
Establishment of Sister Chromatid Cohesion	11	11	5	0.4545455	0.133333333	10	11	3	0	6.79E-08	1.77E-05	0.378012	0.01588
cell division	14	14	5	0.3571429	0.133333333	10	11	3	0	4.16E-07	7.24E-05	0.164799	0.0071
Tie2 Signaling	18	18	5	0.2777778	0.08	6	6	2	0	2.38E-06	0.0003101	0.89513	0.00258
Axon guidance	274	272	18	0.0661765	0.213333333	16	19	2	0	2.00E-05	0.0020847	0	0
Depolarization of the Presynaptic Terminal Triggers the Opening of Calcium Channels	12	11	3	0.2727273	0.04	3	3	1	0	8.85E-05	0.0068915	0.912528	0.00508
GRB2 events in EGFR signaling	22	21	4	0.1904762	0.066666667	5	5	2	0	0.0001056	0.0068915	0.012371	0.00085
SOS-mediated signalling	22	21	4	0.1904762	0.066666667	5	5	2	0	0.0001056	0.0068915	0.013039	0.00056
SHC1 events in EGFR signaling	23	22	4	0.1818182	0.066666667	5	5	2	0	0.0001341	0.0069992	0.012481	0.00032
SHC-mediated signalling	24	22	4	0.1818182	0.066666667	5	5	2	0	0.0001341	0.0069992	0.012396	0.0007
Integrin cell surface interactions	86	85	8	0.0941176	0.106666667	8	8	1	0	0.0001665	0.0072421	0.58028	0.01986
Developmental Biology	450	427	22	0.0515222	0.32	24	28	6	0	0.0001588	0.0072421	0	0
SHC-related events	26	24	4	0.1666667	0.066666667	5	5	2	0	0.0002083	0.0083631	0.008592	0.00085
SHC-related events triggered by IGF1R	30	25	4	0.16	0.066666667	5	5	2	0	0.0002554	0.009523	0.006948	0.00068
Interaction between L1 and Ankyrins	26	26	4	0.1538462	0.053333333	4	4	1	0	0.0003102	0.010121	0.866428	0.01072

Cell surface interactions at the vascular wall	93	92	8	0.0869565	0.12	9	9	2	0	0.0003051	0.010121	0.000738	0.00022
Activation of Na-permeable Kainate Receptors	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.074249	0.00415
Conjugation of phenylacetate with glutamine	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.096762	0.0052
Conjugation of salicylate with glycine	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.09714	0.00287
ERK2 activation	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.063926	0.0039
phospho-PLA2 pathway	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.068015	0.00362
Transcriptional activation of cell cycle inhibitor p21	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.106041	0.00564
Transcriptional activation of p53 responsive genes	2	2	1	0.5	0.013333333	1	1	1	0	0.0005602	0.0102789	0.10818	0.00704
Spry regulation of FGF signaling	16	15	3	0.2	0.04	3	3	1	0	0.0003399	0.0102789	0.143883	0.00822
Regulation of KIT signaling	16	16	3	0.1875	0.04	3	3	1	0	0.0004448	0.0102789	0.223542	0.00962
RAF/MAP kinase cascade	18	17	3	0.1764706	0.053333333	4	4	2	0	0.000571	0.0102789	0.016662	0.00121
SHC1 events in ERBB4 signaling	31	28	4	0.1428571	0.066666667	5	5	2	0	0.0004461	0.0102789	0.003914	0.00039
NCAM signaling for neurite out-growth	78	77	7	0.0909091	0.106666667	8	8	2	0	0.0004402	0.0102789	0.008468	0.00052
L1CAM interactions	95	94	8	0.0851064	0.08	6	8	1	0	0.0003588	0.0102789	0.038617	0.0042
GRB2 events in ERBB2 signaling	33	30	4	0.1333333	0.066666667	5	5	2	0	0.0006226	0.0108329	0.00171	0.00024
Signaling by constitutively active EGFR	19	18	3	0.1666667	0.053333333	4	4	2	0	0.0007207	0.0121364	0.01973	0.00119
SHC1 events in ERBB2 signaling	36	32	4	0.125	0.066666667	5	5	2	0	0.0008468	0.0138136	0.001755	0.00016

EGFR Transactivation by Gastrin	9	9	2	0.2222222	0.04	3	3	2	0	0.0009924	0.0156981	0.005447	0.00052
Signalling to RAS	40	34	4	0.1176471	0.066666667	5	5	2	0	0.0011263	0.017292	0.001002	0.00013
Netrin mediated repulsion signals	10	10	2	0.2	0.026666667	2	2	1	0	0.0013931	0.0207777	0.05314	0.00116
Amino Acid conjugation	3	3	1	0.3333333	0.013333333	1	1	1	0	0.0016544	0.0221436	0.076658	0.00241
Conjugation of benzoate with glycine	3	3	1	0.3333333	0.013333333	1	1	1	0	0.0016544	0.0221436	0.081181	0.00226
Conjugation of carboxylic acids	3	3	1	0.3333333	0.013333333	1	1	1	0	0.0016544	0.0221436	0.075189	0.00354
Signalling to p38 via RIT and RIN	23	22	3	0.1363636	0.053333333	4	4	2	0	0.0016003	0.0221436	0.005257	0.00022
Interleukin-6 signaling	13	11	2	0.1818182	0.026666667	2	2	1	0	0.0018824	0.0245651	0.024076	0.0013
Activation of Ca-permeable Kainate Receptor	10	12	2	0.1666667	0.026666667	2	2	1	0	0.0024664	0.0268222	0.220422	0.00738
ERK activation	13	12	2	0.1666667	0.026666667	2	2	1	0	0.0024664	0.0268222	0.012922	0.00048
Ionotropic activity of Kainate Receptors	10	12	2	0.1666667	0.026666667	2	2	1	0	0.0024664	0.0268222	0.216852	0.00627
Metabolism of Angiotensinogen to Angiotensins	12	12	2	0.1666667	0.026666667	2	2	1	0	0.0024664	0.0268222	0.307955	0.01597
ARMS-mediated activation	26	24	3	0.125	0.053333333	4	4	2	0	0.0022405	0.0268222	0.006005	0.00063
Lysosome Vesicle Biogenesis	24	24	3	0.125	0.04	3	3	1	0	0.0022405	0.0268222	0.349359	0.01275
S Phase	125	121	8	0.0661157	0.173333333	13	14	3	0	0.0022151	0.0268222	0	0
Hemostasis	454	457	20	0.0437637	0.253333333	19	23	3	0	0.0024465	0.0268222	0	0
Frs2-mediated activation	28	25	3	0.12	0.053333333	4	4	2	0	0.0026186	0.0278959	0.004132	0.00046
Activation of NOXA and translocation to mitochondria	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.064183	0.00277

Activation of PUMA and translocation to mitochondria	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.069191	0.00149
Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.077643	0.00306
G2/M DNA replication checkpoint	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.078058	0.00297
LDL endocytosis	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.099038	0.00598
Polo-like kinase mediated events	4	4	1	0.25	0.013333333	1	1	1	0	0.0032572	0.0295497	0.078251	0.00381
Transport of vitamins, nucleosides, and related molecules	13	13	2	0.1538462	0.026666667	2	2	1	0	0.0031509	0.0295497	0.242654	0.01326
Prolonged ERK activation events	29	26	3	0.1153846	0.053333333	4	4	2	0	0.0030383	0.0295497	0.003865	0.00028
Signalling to ERKs	51	43	4	0.0930233	0.066666667	5	5	2	0	0.0032833	0.0295497	0.00027	#####
FRS2-mediated cascade	50	44	4	0.0909091	0.066666667	5	5	2	0	0.0036346	0.0321572	0.000154	#####
Recycling pathway of L1	29	28	3	0.1071429	0.04	3	3	1	0	0.004011	0.0348957	0.103943	0.00519
Activation of RAS in B Cells	5	5	1	0.2	0.026666667	2	2	2	0	0.0053442	0.0404301	0.009004	0.00051
Ca activated K+ channels	5	5	1	0.2	0.013333333	1	1	1	0	0.0053442	0.0404301	0.07852	0.00199
NOSTRIN mediated eNOS trafficking	6	5	1	0.2	0.013333333	1	1	1	0	0.0053442	0.0404301	0.034371	0.00219
RAF activation	5	5	1	0.2	0.026666667	2	2	2	0	0.0053442	0.0404301	0.007785	0.00044
Retinoid metabolism and transport	5	5	1	0.2	0.013333333	1	1	1	0	0.0053442	0.0404301	0.0777	0.00415
Serotonin and melatonin biosynthesis	7	5	1	0.2	0.013333333	1	1	1	0	0.0053442	0.0404301	0.080303	0.00355

GRB2:SOS provides linkage to MAPK signaling for Intergrins	15	15	2	0.1333333	0.026666667	2	2	1	0	0.0048413	0.0404301	0.013047	0.00101
Interleukin-2 signaling	48	47	4	0.0851064	0.066666667	5	5	2	0	0.00485	0.0404301	0.000152	#####
Neuronal System	301	272	13	0.0477941	0.173333333	13	13	1	0	0.0049965	0.0404301	0	0
Other semaphorin interactions	16	16	2	0.125	0.026666667	2	2	1	0	0.0058558	0.0436678	0.211202	0.00962
Platelet Aggregation (Plug Formation)	35	32	3	0.09375	0.04	3	3	1	0	0.0065464	0.0481299	0.04971	0.00277

path id	full path length	testable path length	aberrational genes	proportion aberrational	proportion of cohort w aberrational gene in path	Num patient with aberrational gene (s) in path	count of aberrational genes in path cross cohort	max in one gene	min in one gene	hyperg p value	hyperg p w FDR
Cohesin Loading onto Chromatin	10	10	5	0.5	0.13333333	10	11	3	0	3.15E-08	1.64E-05
Activation of Na-permeable Kainate Receptors	2	2	1	0.5	0.01333333	1	1	1	0	0.000560221	0.010278856
Conjugation of phenylacetate with glutamine	2	2	1	0.5	0.01333333	1	1	1	0	0.000560221	0.010278856
Conjugation of salicylate with glycine	2	2	1	0.5	0.01333333	1	1	1	0	0.000560221	0.010278856
Transcriptional activation of cell cycle inhibitor p21	2	2	1	0.5	0.01333333	1	1	1	0	0.000560221	0.010278856
Transcriptional activation of p53 responsive genes	2	2	1	0.5	0.01333333	1	1	1	0	0.000560221	0.010278856
Establishment of Sister Chromatid Cohesion	11	11	5	0.454545455	0.13333333	10	11	3	0	6.79E-08	1.77E-05
Amino Acid conjugation	3	3	1	0.333333333	0.01333333	1	1	1	0	0.001654407	0.022143602
Conjugation of benzoate with glycine	3	3	1	0.333333333	0.01333333	1	1	1	0	0.001654407	0.022143602
Conjugation of carboxylic acids	3	3	1	0.333333333	0.01333333	1	1	1	0	0.001654407	0.022143602

Depolarization of the Presynaptic Terminal Triggers the Opening of Calcium Channels	12	11	3	0.272727273	0.04	3	3	1	0	8.85E-05	0.006891504
Activation of NOXA and translocation to mitochondria	4	4	1	0.25	0.01333333	1	1	1	0	0.003257221	0.029549671
Activation of PUMA and translocation to mitochondria	4	4	1	0.25	0.01333333	1	1	1	0	0.003257221	0.029549671
LDL endocytosis	4	4	1	0.25	0.01333333	1	1	1	0	0.003257221	0.029549671
Activation of RAS in B Cells	5	5	1	0.2	0.02666667	2	2	2	0	0.005344214	0.040430139
Ca activated K+ channels	5	5	1	0.2	0.01333333	1	1	1	0	0.005344214	0.040430139
NOSTRIN mediated eNOS trafficking	6	5	1	0.2	0.01333333	1	1	1	0	0.005344214	0.040430139
Retinoid metabolism and transport	5	5	1	0.2	0.01333333	1	1	1	0	0.005344214	0.040430139
Serotonin and melatonin biosynthesis	7	5	1	0.2	0.01333333	1	1	1	0	0.005344214	0.040430139
Activation of Ca-permeable Kainate Receptor	10	12	2	0.166666667	0.02666667	2	2	1	0	0.002466405	0.026822157
Ionotropic activity of Kainate Receptors	10	12	2	0.166666667	0.02666667	2	2	1	0	0.002466405	0.026822157

Metabolism of Angiotensinogen to Angiotensins	12	12	2	0.166666667	0.02666667	2	2	1	0	0.002466405	0.026822157
Interaction between L1 and Ankyrins	26	26	4	0.153846154	0.05333333	4	4	1	0	0.000310222	0.010121001
Transport of vitamins, nucleosides, and related molecules	13	13	2	0.153846154	0.02666667	2	2	1	0	0.0031509	0.029549671
Lysosome Vesicle Biogenesis	24	24	3	0.125	0.04	3	3	1	0	0.002240476	0.026822157
Other semaphorin interactions	16	16	2	0.125	0.02666667	2	2	1	0	0.005855829	0.043667754

Sensitive and Cryptic Pathways. Showing top 104 pathways, sorted by hypergeometric p-value for sensitive gene enrichment

path id	F P L	T P L drg Scn	# D S Gen	P P D S	P. co. w D sens in pth	# pat W D S G in Pth	C D S x in Pths x coh	Mx O G	Mi O G	Hyp G	Hyp G FDR	T P L Ab	ab. gen	p. ab.	P. co. w ab in pth	# pat w ab in pth	# ab x co	Mx O G	Mi O G	Hyp G	Hyp G FDR
Role of Abl in Robo-Slit signaling	9	2	2	1	0.2059	7	9	6	3	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
VEGF binds to VEGFR leading to receptor dimerization	17	3	3	1	0.0882	3	5	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Downregulation of ERBB4 signaling	10	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G2 Phase	5	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DNA Replication Pre-Initiation	83	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
M/G1 Transition	83	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of DNA replication	76	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Removal of licensing factors from origins	73	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Orc1 removal from chromatin	71	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Switching of origins to a post-replicative state	71	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Regulation of gap junction activity	4	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ADP signalling through P2Y purinoceptor 1	24	1	1	1	0.1765	6	6	6	6	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SCF(Skp2)-mediated degradation of p27/p21	56	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

CDK-mediated phosphorylation and removal of Cdc6	50	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Activation of the pre-replicative complex	31	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G0 and Early G1	25	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Nef Mediated CD4 Down-regulation	36	1	1	1	0.0882	3	3	3	3	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GRB7 events in ERBB2 signaling	6	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NRIF signals cell death from the nucleus	27	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Activation of BMF and translocation to mitochondria	3	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Platelet Adhesion to exposed collagen	13	2	2	1	0.1176	4	6	4	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Signaling by VEGF	22	3	3	1	0.0882	3	5	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
VEGF ligand-receptor interactions	22	3	3	1	0.0882	3	5	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Interleukin-7 signaling	9	2	2	1	0.0588	2	3	2	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Neurophilin interactions with VEGF and VEGFR	9	2	2	1	0.0882	3	4	3	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Phosphorylation of proteins involved in G1/S transition by active Cyclin E:Cdk2 complexes	4	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Activation of BIM and translocation to mitochondria	3	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
c-src mediated regulation of Cx43 function and closure of gap junctions	4	1	1	1	0.0294	1	1	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DNA Replication	104	1	1	1	0.0294	1	1	1	1	0	0	102	1	0.0098	0.01333	1	1	1	0	0.70257	0.74089
Synthesis of DNA	96	1	1	1	0.0294	1	1	1	1	0	0	94	1	0.0106	0.01333	1	1	1	0	0.65902	0.6992
Meiotic Recombination	54	1	1	1	0.0294	1	1	1	1	0	0	84	1	0.0119	0.01333	1	1	1	0	0.59766	0.64193
p53-Dependent G1 DNA Damage Response	57	1	1	1	0.0294	1	1	1	1	0	0	56	1	0.0179	0.01333	1	1	1	0	0.38561	0.46595
p53-Dependent G1/S DNA damage checkpoint	57	1	1	1	0.0294	1	1	1	1	0	0	56	1	0.0179	0.01333	1	1	1	0	0.38561	0.46595
Nuclear signaling by ERBB4	50	2	2	1	0.0588	2	3	2	1	0	0	41	1	0.0244	0.01333	1	1	1	0	0.25436	0.35033
Interferon alpha/beta signaling	113	2	2	1	0.0294	1	2	1	1	0	0	65	2	0.0308	0.02667	2	2	1	0	0.20011	0.29424
Cell death signalling via NRAGE, NRIF and NADE	119	1	1	1	0.0294	1	1	1	1	0	0	63	2	0.0317	0.02667	2	2	1	0	0.18795	0.28112
Signal amplification	30	1	1	1	0.1765	6	6	6	6	0	0	30	1	0.0333	0.01333	1	1	1	0	0.15863	0.25558
EGFR downregulation	31	1	1	1	0.0588	2	2	2	2	0	0	26	1	0.0385	0.01333	1	1	1	0	0.12583	0.22806
Regulation of IFNA signaling	26	2	2	1	0.0294	1	2	1	1	0	0	26	1	0.0385	0.01333	1	1	1	0	0.12583	0.22806
NRAGE signals death through JNK	90	1	1	1	0.0294	1	1	1	1	0	0	47	2	0.0426	0.02667	2	2	1	0	0.09996	0.19543

Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters	48	1	1	1	0.0882	3	3	3	3	0	0	20	1	0.05	0.01333	1	1	1	0	0.08054	0.17404
Platelet sensitization by LDL	22	2	2	1	0.2353	8	9	6	3	0	0	17	1	0.0588	0.01333	1	1	1	0	0.06035	0.15038
Meiosis	85	1	1	1	0.0294	1	1	1	1	0	0	115	5	0.0435	0.14667	11	11	3	0	0.05528	0.14499
Prolactin receptor signaling	15	1	1	1	0.0588	2	2	2	2	0	0	16	1	0.0625	0.01333	1	1	1	0	0.05407	0.14256
Gap junction trafficking and regulation	22	1	1	1	0.0294	1	1	1	1	0	0	14	1	0.0714	0.01333	1	1	1	0	0.04229	0.12834
Regulation of IFNG signaling	14	2	2	1	0.0588	2	3	2	1	0	0	14	1	0.0714	0.01333	1	1	1	0	0.04229	0.12834
PECAM1 interactions	12	5	5	1	0.1176	4	12	4	1	0	0	12	1	0.0833	0.01333	1	1	1	0	0.03163	0.11794
Golgi Associated Vesicle Biogenesis	49	1	1	1	0.0294	1	1	1	1	0	0	49	3	0.0612	0.04	3	3	1	0	0.02832	0.11284
Clathrin derived vesicle budding	56	1	1	1	0.0294	1	1	1	1	0	0	56	4	0.0714	0.05333	4	4	1	0	0.01019	0.0638
trans-Golgi Network Vesicle Budding	56	1	1	1	0.0294	1	1	1	1	0	0	56	4	0.0714	0.05333	4	4	1	0	0.01019	0.0638
DSCAM interactions	14	5	4	0.8	0.1765	6	12	6	0	0.0004	0.0017	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
p38MAPK events	17	5	4	0.8	0.1765	6	12	6	0	0.0004	0.0017	13	1	0.0769	0.02667	2	2	2	0	0.03681	0.12317
CDO in myogenesis	30	5	4	0.8	0.2353	8	17	6	0	0.0004	0.0017	30	2	0.0667	0.02667	2	2	1	0	0.03336	0.11845
Myogenesis	30	5	4	0.8	0.2353	8	17	6	0	0.0004	0.0017	30	2	0.0667	0.02667	2	2	1	0	0.03336	0.11845
Interleukin-3, 5 and GM-CSF signaling	46	13	7	0.538	0.2059	7	15	4	0	0.0014	0.0061	45	3	0.0667	0.04	3	3	1	0	0.02141	0.1001
Nef and signal transduction	35	4	3	0.75	0.1176	4	7	3	0	0.002	0.0082	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

The role of Nef in HIV-1 replication and disease pathogenesis	55	4	3	0.75	0.1176	4	7	3	0	0.002	0.0082	27	1	0.037	0.01333	1	1	1	0	0.13387	0.23769
Antigen Activates B Cell Receptor Leading to Generation of Second Messengers	171	6	4	0.667	0.1471	5	9	4	0	0.002	0.0082	29	2	0.069	0.02667	2	2	1	0	0.03054	0.11794
B Cell Activation	344	26	11	0.423	0.2647	9	34	7	0	0.0022	0.0088	190	5	0.0263	0.08	6	6	2	0	0.29672	0.39715
Signaling by SCF-KIT	155	37	14	0.378	0.2647	9	38	7	0	0.0026	0.0098	142	7	0.0493	0.10667	8	8	2	0	0.01933	0.09608
CTLA4 inhibitory signaling	28	9	5	0.556	0.1176	4	12	4	0	0.0038	0.0139	22	1	0.0455	0.01333	1	1	1	0	0.095	0.18713
Immune System	1335	77	23	0.299	0.3824	13	59	7	0	0.0048	0.0171	1040	17	0.0163	0.2	15	18	2	0	0.95071	0.95711
Cell-Cell communication	132	12	6	0.5	0.1765	6	15	6	0	0.0048	0.0171	128	6	0.0469	0.09333	7	7	2	0	0.03213	0.11845
Activation of the AP-1 family of transcription factors	10	7	4	0.571	0.1765	6	13	6	0	0.0059	0.0202	10	1	0.1	0.01333	1	1	1	0	0.02224	0.1001
Signaling by ERBB4	175	29	11	0.379	0.2647	9	31	7	0	0.0072	0.0243	153	5	0.0327	0.08	6	6	2	0	0.15576	0.25558
Constitutive PI3K/AKT Signaling in Cancer	94	19	8	0.421	0.2353	8	27	7	0	0.0078	0.0261	84	2	0.0238	0.02667	2	2	1	0	0.32184	0.41687
Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	53	5	3	0.6	0.1765	6	11	6	0	0.0084	0.0271	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NOD1/2 Signaling Pathway	31	5	3	0.6	0.1765	6	11	6	0	0.0084	0.0271	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Host Interactions of HIV factors	318	5	3	0.6	0.1176	4	7	3	0	0.0084	0.0271	126	2	0.0159	0.32	24	24	23	0	0.58002	0.62815
Signaling by ERBB2	171	37	13	0.351	0.2647	9	34	7	0	0.0088	0.0279	159	5	0.0314	0.08	6	6	2	0	0.17632	0.26912
DCC mediated attractive signaling	14	3	2	0.667	0.0588	2	3	2	0	0.0098	0.03	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Destabilization of mRNA by KSRP	17	3	2	0.667	0.1765	6	10	6	0	0.0098	0.03	17	1	0.0588	0.02667	2	2	2	0	0.06035	0.15038
GAB1 signalosome	113	23	9	0.391	0.2353	8	28	7	0	0.0103	0.031	102	2	0.0196	0.02667	2	2	1	0	0.43829	0.50395
Signaling by NGF	374	50	16	0.32	0.3235	11	44	7	0	0.012	0.0357	285	7	0.0246	0.10667	8	8	2	0	0.36542	0.45309
Regulation of signaling by CBL	19	8	4	0.5	0.1471	5	10	4	0	0.0132	0.0387	18	1	0.0556	0.01333	1	1	1	0	0.06687	0.16387
Adaptive Immune System	856	39	13	0.333	0.2647	9	37	7	0	0.0156	0.0453	606	13	0.0215	0.16	12	14	2	0	0.58549	0.63276
PI3K/AKT Signaling in Cancer	109	21	8	0.381	0.2353	8	27	7	0	0.0175	0.0481	98	2	0.0204	0.02667	2	2	1	0	0.41287	0.48322
PI-3K cascade	109	21	8	0.381	0.2353	8	27	7	0	0.0175	0.0481	98	2	0.0204	0.02667	2	2	1	0	0.41287	0.48322
PI3K events in ERBB2 signaling	109	21	8	0.381	0.2353	8	27	7	0	0.0175	0.0481	98	2	0.0204	0.02667	2	2	1	0	0.41287	0.48322
PI3K events in ERBB4 signaling	109	21	8	0.381	0.2353	8	27	7	0	0.0175	0.0481	98	2	0.0204	0.02667	2	2	1	0	0.41287	0.48322
PIP3 activates AKT signaling	109	21	8	0.381	0.2353	8	27	7	0	0.0175	0.0481	98	2	0.0204	0.02667	2	2	1	0	0.41287	0.48322
Innate Immune System	649	56	17	0.304	0.3529	12	47	7	0	0.0186	0.0503	445	7	0.0157	0.10667	8	8	2	0	0.83972	0.86457
Downstream signal transduction	170	36	12	0.333	0.2647	9	32	7	0	0.0198	0.0525	158	5	0.0316	0.08	6	6	2	0	0.17282	0.26912
Signaling by PDGF	199	36	12	0.333	0.2647	9	32	7	0	0.0198	0.0525	187	7	0.0374	0.10667	8	8	2	0	0.07623	0.17168
NGF signalling via TRKA from the plasma membrane	225	48	15	0.313	0.3235	11	43	7	0	0.0201	0.0528	202	6	0.0297	0.09333	7	7	2	0	0.20386	0.29792
Signaling by Interleukins	112	25	9	0.36	0.2059	7	19	4	0	0.0208	0.0539	110	5	0.0455	0.08	6	6	2	0	0.04625	0.13341

HIV Infection	424	6	3	0.5	0.1176	4	7	3	0	0.0213	0.0541	198	4	0.0202	0.33333	25	26	23	0	0.5094	0.55981
Interleukin receptor SHC signaling	28	6	3	0.5	0.0588	2	5	2	0	0.0213	0.0541	28	1	0.0357	0.01333	1	1	1	0	0.14202	0.24387
CD28 co-stimulation	33	12	5	0.417	0.1176	4	12	4	0	0.0247	0.0592	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Downstream Signaling Events Of B Cell Receptor (BCR)	179	22	8	0.364	0.2353	8	27	7	0	0.0249	0.0592	167	3	0.018	0.05333	4	4	2	0	0.5647	0.61284
PI3K/AKT activation	111	22	8	0.364	0.2353	8	27	7	0	0.0249	0.0592	100	2	0.02	0.02667	2	2	1	0	0.42563	0.49373
DAP12 interactions	185	37	12	0.324	0.2941	10	33	7	0	0.0257	0.0592	170	5	0.0294	0.08	6	6	2	0	0.21666	0.31415
DAP12 signaling	169	37	12	0.324	0.2941	10	33	7	0	0.0257	0.0592	160	5	0.0313	0.08	6	6	2	0	0.17985	0.27291
Signaling by EGFR in Cancer	193	37	12	0.324	0.2647	9	32	7	0	0.0257	0.0592	178	6	0.0337	0.09333	7	7	2	0	0.13014	0.23506
Signaling by EGFR	191	37	12	0.324	0.2647	9	32	7	0	0.0257	0.0592	176	6	0.0341	0.09333	7	7	2	0	0.12471	0.22806
Regulation of mitotic cell cycle	82	4	2	0.5	0.0588	2	2	1	0	0.0333	0.0713	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CD28 dependent Vav1 pathway	11	4	2	0.5	0.0882	3	5	3	0	0.0333	0.0713	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
APC/C-mediated degradation of cell cycle proteins	82	4	2	0.5	0.0588	2	2	1	0	0.0333	0.0713	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Platelet homeostasis	87	4	2	0.5	0.2353	8	9	6	0	0.0333	0.0713	81	2	0.0247	0.02667	2	2	1	0	0.3022	0.40345
Growth hormone receptor signaling	24	4	2	0.5	0.1765	6	6	4	0	0.0333	0.0713	25	1	0.04	0.01333	1	1	1	0	0.11791	0.2214
Signaling by FGFR	171	38	12	0.316	0.2647	9	32	7	0	0.033	0.0713	158	7	0.0443	0.10667	8	8	2	0	0.03386	0.11943
Signaling by FGFR in disease	190	38	12	0.316	0.2647	9	32	7	0	0.033	0.0713	173	8	0.0462	0.12	9	9	2	0	0.02151	0.1001
Netrin-1 signaling	41	4	2	0.5	0.0588	2	3	2	0	0.0333	0.0713	42	3	0.0714	0.04	3	3	1	0	0.01698	0.09138
Phosphorylation of CD3 and TCR zeta chains	39	2	1	0.5	0.0882	3	3	3	0	0.047	0.0802	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Aberration Enriched Pathways. Showing top 81 pathways, sorted by number of aberrational genes

path id	full path length	testable path length	mutated genes	proportion mutated	proportion of cohort w mutated gene in path	Num patient with mutated gene (s) in path	count of mutated genes in path cross cohort	max in one gene	min in one gene	hyperg p value	hyperg p w FDR	mgsa probability estimate	mgsa std error
Signaling Pathways	1874	1653	1043	0.631	0.987616	319	2898	222	0	0.00318	0.034144	0.999	3E-04
Developmental Biology	450	427	318	0.7447	0.900929	291	1141	56	0	7.03E-11	1.24E-09	0.997	8E-04
Transmembrane transport of small molecules	425	410	291	0.7098	0.798762	258	738	12	0	1.30E-06	2.18E-05	0.998	6E-04
Generic Transcription Pathway	496	430	284	0.6605	0.786378	254	733	56	0	0.00486	0.044987	0.212	0.197
Axon guidance	274	272	213	0.7831	0.817337	264	825	16	0	3.69E-11	6.63E-10	0.019	0.001
Neuronal System	301	272	202	0.7426	0.678019	219	570	13	0	2.83E-07	4.87E-06	0.999	2E-04
Signaling by NGF	374	285	199	0.6982	0.758514	245	604	48	0	0.00026	0.003485	0.016	5E-04
SLC-mediated transmembrane transport	230	226	162	0.7168	0.591331	191	351	11	0	0.00011	0.001593	0.017	6E-04
NGF signalling via TRKA from the plasma membrane	225	202	141	0.698	0.653251	211	415	48	0	0.00177	0.02104	0.016	4E-04
Signaling by PDGF	199	187	138	0.738	0.73065	236	476	48	0	3.15E-05	0.00048	0.017	9E-04
Transmission across Chemical Synapses	189	180	131	0.7278	0.582043	188	384	13	0	0.00014	0.002044	0.016	4E-04

Signaling by EGFR in Cancer	193	178	124	0.6966	0.640867	207	370	48	0	0.00345	0.03582	0.016	8E-04
Signaling by EGFR	191	176	122	0.6932	0.637771	206	365	48	0	0.00481	0.044936	0.016	7E-04
Signaling by ERBB2	171	159	112	0.7044	0.613003	198	350	48	0	0.00281	0.030997	0.017	6E-04
Extracellular matrix organization	147	145	111	0.7655	0.634675	205	440	20	0	9.81E-06	0.000153	0.998	5E-04
Downstream signal transduction	170	158	111	0.7025	0.634675	205	357	48	0	0.00336	0.035804	0.016	9E-04
Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell	140	133	100	0.7519	0.510836	165	298	12	0	9.18E-05	0.001365	0.017	4E-04
Rho GTPase cycle	212	122	96	0.7869	0.535604	173	309	19	0	4.11E-06	6.69E-05	0.016	2E-04
Signaling by Rho GTPases	212	122	96	0.7869	0.535604	173	309	19	0	4.11E-06	6.69E-05	0.016	5E-04
Regulation of Lipid Metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	120	113	81	0.7168	0.433437	140	210	20	0	0.00417	0.039437	0.016	4E-04
L1CAM interactions	95	94	79	0.8404	0.544892	176	316	16	0	1.28E-07	2.23E-06	0.016	6E-04

PPARA Activates Gene Expression	115	110	79	0.7182	0.414861	134	199	20	0	0.00419	0.039437	0.015	3E-04
Collagen formation	86	86	69	0.8023	0.513932	166	307	20	0	2.09E-05	0.000323	0.017	6E-04
Integrin cell surface interactions	86	85	63	0.7412	0.458204	148	233	22	0	0.00255	0.028607	0.016	2E-04
NCAM signaling for neurite out-growth	78	77	60	0.7792	0.495356	160	261	12	0	0.00031	0.004178	0.016	4E-04
Potassium Channels	79	78	59	0.7564	0.328173	106	161	12	0	0.0014	0.017049	0.016	8E-04
Collagen biosynthesis and modifying enzymes	64	64	56	0.875	0.452012	146	243	20	0	3.21E-07	5.44E-06	0.016	7E-04
Semaphorin interactions	66	66	51	0.7727	0.408669	132	206	12	0	0.0011	0.013894	0.016	8E-04
Ion channel transport	56	56	46	0.8214	0.343653	111	156	8	0	0.00012	0.001651	0.016	6E-04
GABA receptor activation	50	50	39	0.78	0.306502	99	123	12	0	0.00245	0.028043	0.016	5E-04
NRAGE signals death through JNK	90	47	38	0.8085	0.349845	113	155	19	0	0.00072	0.009482	0.016	7E-04
Voltage gated Potassium channels	43	43	38	0.8837	0.226006	73	107	6	0	9.29E-06	0.000147	0.017	4E-04
NCAM1 interactions	44	44	37	0.8409	0.383901	124	175	12	0	0.00015	0.00216	0.015	3E-04
Interleukin-3, 5 and GM-CSF signaling	46	45	35	0.7778	0.334365	108	127	48	0	0.00402	0.03845	0.017	0.001

G-protein mediated events	41	43	35	0.814	0.287926	93	115	12	0	0.00083	0.010719	0.015	5E-04
PLC beta mediated events	40	42	34	0.8095	0.28483	92	114	12	0	0.00116	0.014589	0.016	6E-04
Hexose uptake	45	42	33	0.7857	0.195046	63	75	6	0	0.0036	0.036427	0.016	7E-04
Elastic fibre formation	35	35	28	0.8	0.247678	80	98	10	0	0.00372	0.036855	0.017	1E-03
Signal transduction by L1	35	35	28	0.8	0.226006	73	87	11	0	0.00372	0.036855	0.016	9E-04
Apoptotic cleavage of cellular proteins	33	32	27	0.8438	0.321981	104	127	29	0	0.00077	0.010106	0.064	0.005
Adherens junctions interactions	29	29	27	0.931	0.260062	84	107	14	0	8.34E-06	0.000134	0.069	0.007
PLCG1 events in ERBB2 signaling	31	33	27	0.8182	0.247678	80	97	12	0	0.00216	0.025466	0.016	6E-04
GPVI-mediated activation cascade	33	32	26	0.8125	0.312693	101	113	48	0	0.00307	0.033335	0.016	3E-04
EGFR interacts with phospholipase C-gamma	30	32	26	0.8125	0.241486	78	93	12	0	0.00307	0.033335	0.016	6E-04
Interaction between L1 and Ankyrins	26	26	24	0.9231	0.303406	98	141	16	0	3.45E-05	0.000519	0.016	3E-04
Ion transport by P-type ATPases	28	28	23	0.8214	0.22291	72	85	8	0	0.00347	0.03582	0.016	6E-04

Recycling pathway of L1	29	28	23	0.8214	0.157895	51	58	6	0	0.00347	0.03582	0.015	7E-04
RORA Activates Circadian Expression	26	26	22	0.8462	0.229102	74	88	20	0	0.00173	0.020849	0.017	4E-04
Synthesis of IP3 and IP4 in the cytosol	25	27	22	0.8148	0.20743	67	77	10	0	0.005	0.045934	0.036	0.02
Circadian Repression of Expression by REV-ERBA	25	25	21	0.84	0.216718	70	83	20	0	0.00257	0.028607	0.016	5E-04
Nephrin interactions	22	22	20	0.9091	0.303406	98	112	48	0	0.00022	0.003075	0.037	0.002
Molecules associated with elastic fibres	24	24	20	0.8333	0.195046	63	68	10	0	0.00379	0.036855	0.016	9E-04
CD28 dependent PI3K/Akt signaling	23	21	18	0.8571	0.229102	74	79	48	0	0.00256	0.028607	0.016	6E-04
Regulation of signaling by CBL	19	18	17	0.9444	0.232198	75	82	48	0	0.00011	0.001597	0.016	5E-04
Platelet calcium homeostasis	19	19	17	0.8947	0.173375	56	61	8	0	0.0009	0.011507	0.016	5E-04
PKA activation in glucagon signalling	17	17	16	0.9412	0.160991	52	58	12	0	0.00018	0.00253	0.016	6E-04
Ligand-gated ion channel transport	17	17	15	0.8824	0.142415	46	51	6	0	0.00224	0.02613	0.016	6E-04
Other semaphorin interactions	16	16	14	0.875	0.176471	57	68	12	0	0.00351	0.03582	0.016	5E-04

CRMPs in Sema3A signaling	16	16	14	0.875	0.133127	43	48	9	0	0.00351	0.03582	0.016	6E-04
Sema3A PAK dependent Axon repulsion	15	15	13	0.8667	0.164087	53	58	9	0	0.00549	0.049743	0.015	3E-04
p130Cas linkage to MAPK signaling for integrins	15	15	13	0.8667	0.123839	40	43	10	0	0.00549	0.049743	0.016	7E-04
Caspase-mediated cleavage of cytoskeletal proteins	12	12	12	1	0.22291	72	77	29	1	0	0	0.018	4E-04
Adenylate cyclase inhibitory pathway	13	13	12	0.9231	0.142415	46	50	12	0	0.00139	0.017049	0.016	6E-04
Inhibition of adenylate cyclase pathway	13	13	12	0.9231	0.142415	46	50	12	0	0.00139	0.017049	0.016	4E-04
GABA A receptor activation	12	12	12	1	0.126935	41	44	6	1	0	0	0.016	4E-04
Downregulation of ERBB2:ERBB3 signaling	15	12	11	0.9167	0.083591	27	27	5	0	0.00231	0.02669	0.016	9E-04
Adenylate cyclase activating pathway	10	10	10	1	0.139319	45	48	12	1	0	0	0.016	6E-04

Cohesin Loading onto Chromatin	10	10	10	1	0.083591	27	27	6	1	0	0	0.016	4E-04
Recycling of bile acids and salts	11	11	10	0.9091	0.083591	27	27	6	0	0.00383	0.036855	0.016	4E-04
Synthesis of IP2, IP, and Ins in the cytosol	11	11	10	0.9091	0.074303	24	25	6	0	0.00383	0.036855	0.02	0.005
Establishment of Sister Chromatid Cohesion	11	11	10	0.9091	0.068111	22	22	5	0	0.00383	0.036855	0.016	6E-04
CHL1 interactions	8	8	8	1	0.095975	31	32	9	1	0	0	0.015	6E-04
Nef and signal transduction	35	8	8	1	0.083591	27	28	9	1	0	0	0.021	0.002
Vitamin D (calciferol) metabolism	7	7	7	1	0.111455	36	40	15	1	0	0	0.015	4E-04
Downregulation of ERBB4 signaling	10	7	7	1	0.086687	28	29	14	1	0	0	0.016	1E-04
Terminal pathway of complement	7	7	7	1	0.077399	25	26	10	1	0	0	0.016	6E-04
Cation-coupled Chloride cotransporters	7	7	7	1	0.04644	15	16	5	1	0	0	0.016	4E-04
Neurofascin interactions	7	7	7	1	0.04644	15	16	5	1	0	0	0.017	7E-04
Mineralocorticoid biosynthesis	7	7	7	1	0.021672	7	8	2	1	0	0	0.016	4E-04
RSK activation	6	6	6	1	0.049536	16	17	4	1	0	0	0.016	5E-04

path id	full path length	testable path length	aberrational genes	proportion aberrational	proportion of cohort w aberrational gene in path	Num patient with aberrational gene (s) in path	count of aberrational genes in path cross cohort	max in one gene	min in one gene	hyperg p value	hyperg p w FDR
Extracellular matrix organization	147	145	111	0.765517	0.63467492	205	440	20	0	9.81E-06	0.00015
SLC-mediated transmembrane transport	230	226	162	0.716814	0.59133127	191	351	11	0	0.000108	0.00159
Rho GTPase cycle	212	122	96	0.786885	0.53560372	173	309	19	0	4.11E-06	6.69E-05
Signaling by Rho GTPases	212	122	96	0.786885	0.53560372	173	309	19	0	4.11E-06	6.69E-05
Collagen formation	86	86	69	0.802326	0.51393189	166	307	20	0	2.09E-05	0.00032
Collagen biosynthesis and modifying enzymes	64	64	56	0.875	0.45201238	146	243	20	0	3.21E-07	5.44E-06
NCAM1 interactions	44	44	37	0.840909	0.38390093	124	175	12	0	0.000154	0.00216
Ion channel transport	56	56	46	0.821429	0.34365325	111	156	8	0	0.000115	0.00165
Potassium Channels	79	78	59	0.75641	0.32817337	106	161	12	0	0.001402	0.01705
GABA receptor activation	50	50	39	0.78	0.30650155	99	123	12	0	0.002446	0.02804
Interaction between L1 and Ankyrins	26	26	24	0.923077	0.30340557	98	141	16	0	3.45E-05	0.00052
Adherens junctions interactions	29	29	27	0.931034	0.26006192	84	107	14	0	8.34E-06	0.00013
Elastic fibre formation	35	35	28	0.8	0.24767802	80	98	10	0	0.003725	0.03685
RORA Activates Circadian Expression	26	26	22	0.846154	0.22910217	74	88	20	0	0.001732	0.02085
Voltage gated Potassium channels	43	43	38	0.883721	0.22600619	73	107	6	0	9.29E-06	0.00015

Caspase-mediated cleavage of cytoskeletal proteins	12	12	12	1	0.22291022	72	77	29	1	0	0
Ion transport by P-type ATPases	28	28	23	0.821429	0.22291022	72	85	8	0	0.003472	0.03582
Circadian Repression of Expression by REV-ERBA	25	25	21	0.84	0.21671827	70	83	20	0	0.002566	0.02861
Synthesis of IP3 and IP4 in the cytosol	25	27	22	0.814815	0.20743034	67	77	10	0	0.004998	0.04593
Molecules associated with elastic fibres	24	24	20	0.833333	0.19504644	63	68	10	0	0.003785	0.03685
Hexose uptake	45	42	33	0.785714	0.19504644	63	75	6	0	0.0036	0.03643
Other semaphorin interactions	16	16	14	0.875	0.17647059	57	68	12	0	0.003511	0.03582
Platelet calcium homeostasis	19	19	17	0.894737	0.17337461	56	61	8	0	0.000898	0.01151
Adenylate cyclase inhibitory pathway	13	13	12	0.923077	0.14241486	46	50	12	0	0.001389	0.01705
Inhibition of adenylate cyclase pathway	13	13	12	0.923077	0.14241486	46	50	12	0	0.001389	0.01705
Ligand-gated ion channel transport	17	17	15	0.882353	0.14241486	46	51	6	0	0.002235	0.02613
Adenylate cyclase activating pathway	10	10	10	1	0.13931889	45	48	12	1	0	0
GABA A receptor activation	12	12	12	1	0.12693498	41	44	6	1	0	0
Vitamin D (calciferol) metabolism	7	7	7	1	0.11145511	36	40	15	1	0	0
Activation of BID and translocation to mitochondria	4	4	4	1	0.10526316	34	34	29	1	0	0
CHL1 interactions	8	8	8	1	0.09597523	31	32	9	1	0	0

Cohesin Loading onto Chromatin	10	10	10	1	0.08359133	27	27	6	1	0	0
Recycling of bile acids and salts	11	11	10	0.909091	0.08359133	27	27	6	0	0.003826	0.03685
Terminal pathway of complement	7	7	7	1	0.07739938	25	26	10	1	0	0
Synthesis of IP2, IP, and Ins in the cytosol	11	11	10	0.909091	0.07430341	24	25	6	0	0.003826	0.03685
Regulation of Commissural axon pathfinding by Slit and Robo	4	4	4	1	0.07120743	23	23	8	3	0	0
Establishment of Sister Chromatid Cohesion	11	11	10	0.909091	0.06811146	22	22	5	0	0.003826	0.03685
Activation of AMPA receptors	8	4	4	1	0.06501548	21	21	7	3	0	0
ATP sensitive Potassium channels	4	3	3	1	0.05882353	19	19	12	3	0	0
Cation-coupled Chloride cotransporters	7	7	7	1	0.04643963	15	16	5	1	0	0
Neurofascin interactions	7	7	7	1	0.04643963	15	16	5	1	0	0
Serotonin and melatonin biosynthesis	7	5	5	1	0.0371517	12	12	5	1	0	0
Apoptotic cleavage of cell adhesion proteins	6	6	6	1	0.03405573	11	12	3	1	0	0
Class C/3 (Metabotropic glutamate/pheromone receptors)	5	5	5	1	0.02786378	9	9	3	1	0	0

Amino Acid conjugation	3	3	3	1	0.0247678	8	8	4	1	0	0
Conjugation of benzoate with glycine	3	3	3	1	0.0247678	8	8	4	1	0	0
Conjugation of carboxylic acids	3	3	3	1	0.0247678	8	8	4	1	0	0
Lectin pathway of complement activation	3	3	3	1	0.0247678	8	8	3	2	0	0
Mineralocorticoid biosynthesis	7	7	7	1	0.02167183	7	8	2	1	0	0
Astrocytic Glutamate-Glutamine Uptake And Metabolism	4	4	4	1	0.02167183	7	7	3	1	0	0
Neurotransmitter uptake and Metabolism In Glial Cells	4	4	4	1	0.02167183	7	7	3	1	0	0
Conjugation of salicylate with glycine	2	2	2	1	0.02167183	7	7	4	3	0	0
Organic anion transporters	4	4	4	1	0.01857585	6	6	3	1	0	0
Plasmalogen biosynthesis	4	4	4	1	0.01857585	6	6	3	1	0	0
Activation of Na-permeable Kainate Receptors	2	2	2	1	0.01857585	6	6	4	2	0	0
Electric Transmission Across Gap Junctions	4	4	4	1	0.01547988	5	5	2	1	0	0
ER Quality Control Compartment (ERQC)	4	4	4	1	0.01547988	5	5	2	1	0	0
Transmission across Electrical Synapses	4	4	4	1	0.01547988	5	5	2	1	0	0

Beta oxidation of palmitoyl-CoA to myristoyl-CoA	3	3	3	1	0.01547988	5	5	2	1	0	0
Formation of the active cofactor, UDP-glucuronate	3	3	3	1	0.01547988	5	5	2	1	0	0
N-glycan antennae elongation in the medial/trans-Golgi	3	3	3	1	0.01547988	5	5	3	1	0	0
Reactions specific to the complex N-glycan synthesis pathway	3	3	3	1	0.01547988	5	5	3	1	0	0
The NLRP1 inflammasome	3	3	3	1	0.01547988	5	5	3	1	0	0
Conjugation of phenylacetate with glutamine	2	2	2	1	0.01547988	5	5	4	1	0	0
Vitamins	2	2	2	1	0.01547988	5	5	3	2	0	0
vRNP Assembly	5	2	2	1	0.01547988	5	5	4	1	0	0
Synthesis of IPs in the nucleus	4	4	4	1	0.0123839	4	4	1	1	0	0
GABA synthesis	2	2	2	1	0.00928793	3	3	2	1	0	0
Synthesis of PG	2	2	2	1	0.00928793	3	3	2	1	0	0
Synthesis and processing of accessory proteins	28	1	1	1	0.00928793	3	3	3	3	0	0
Entry of Influenza Virion into Host Cell via Endocytosis	11	2	2	1	0.00619195	2	2	1	1	0	0
Nef mediated downregulation of CD28 cell surface expression	29	2	2	1	0.00619195	2	2	1	1	0	0

BoNT Light Chain Types A, C1, E cleave SNAP-25	5	1	1	1	0.00309598	1	1	1	1	0	0
DNA Damage Bypass	2	1	1	1	0.00309598	1	1	1	1	0	0
Translesion synthesis by DNA polymerases bypassing lesion on DNA template	2	1	1	1	0.00309598	1	1	1	1	0	0
Translesion synthesis by Pol zeta	2	1	1	1	0.00309598	1	1	1	1	0	0