Published in final edited form as:

Nat Protoc. 2017 November ; 12(11): 2323–2341. doi:10.1038/nprot.2017.086.

Base-resolution stratification of cancer mutations using functional variomics

Yi Song^{1,*,#}, Liu Ning-Ning^{2,†}, Hu Limei¹, Wang Hui^{2,*}, and Sahni Nidhi^{1,3,*}

¹Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

³Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA

SUMMARY

A complete understanding of human cancer variants requires new methods to systematically and efficiently assess the functional effects of genomic mutations at large scale. Here we describe a set of tools to rapidly clone and stratify thousands of cancer mutations at base resolution. This protocol provides a massively parallel pipeline to achieve high stringency and throughput. The approach includes high-throughput generation of mutant clones by Gateway, confirmation of variant identity by barcoding and next-generation sequencing, as well as stratification of cancer variants by multiplexed interaction profiling. Compared with previous site-directed mutagenesis methods, our protocol requires less sequencing effort and enables robust statistical calling of allele-specific effects. To ensure the precision of variant interaction profiling, we further describe two complementary methods, high-throughput enhanced yeast two-hybrid (HT-eY2H) assay and mammalian cell-based *Gaussia princeps* luciferase protein-fragment complementation assay (GPCA). These independent assays with standard controls validate mutational interaction profiles with high quality. This protocol provides experimentally derived guidelines for classifying candidate cancer alleles emerging from whole-genome or exome sequencing projects as drivers or passengers. For ~100 genomic mutations, the protocol including target primer design, variant library construction and sequence verification can be achieved within as little as 2-3 weeks, and cancer variant stratification can be completed within 2 weeks.

EDITORIAL SUMMARY:

This functional variomics protocol provides a massively parallel pipeline to achieve high-throughput generation of mutant clones by Gateway, confirmation of variant identity by barcoding

*Correspondence should be addressed to Song Yi (Syi2@mdanderson.org), Hui Wang (huiwang@shsmu.edu.cn) and Nidhi Sahni (NSahni@mdanderson.org).

[#]co-first authors

AUTHOR CONTRIBUTIONS

S.Y., L.H. and N.S. performed the experiments. S.Y. and N.S. analyzed data. S.Y. and N.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare that they have no competing financial interests.

and next-generation sequencing, and mechanistic stratification of variants by multiplexed interaction profiling and experimental validation.

Keywords

Functional variomics; Deep mutagenesis; Cancer mutations; high-throughput mutagenesis; barcoded sequencing; Gateway Entry clones; molecular interaction profiling

INTRODUCTION

The ability to stratify and distinguish disease mutations holds enormous potential for systems-level understanding of human cancer, and applications in precision medicine. With rapidly evolving next-generation sequencing platforms, in the past decade there have been hundreds of thousands of human genome sequences available from patients stricken by many cancer types^{1–7}. To evaluate the impact of such a deluge of genotypic information, it is critical to understand the functional and mechanistic molecular consequences of genotypic differences^{8,9}. Accumulating genotypic information in the absence of efficient and informative functional analyses has created a bottleneck in understanding of genotype-phenotype relationships in human cancer¹⁰ (Fig. 1). To this end, a number of deep mutagenesis technologies have emerged in recent years, including PFunkel¹¹, deep mutational scanning^{12,13}, Clone-seq¹⁴ and most recently, Gateway-mediated functional variomics¹⁵. These strategies are all coupled with next-generation sequencing. PFunkel utilizes Pfu DNA polymerase and optimized conditions for PCR-based mutagenesis, and degrades wild-type sequences in the mutant library by adding uracil DNA glycosylase and exonuclease III. Deep mutational scanning constructs a short-sequence mutation library by Gibson assembly and ligation, followed by measurable functional assays. Clone-seq is another PCR-based mutagenesis platform using Phusion polymerase. By contrast, functional variomics integrates Gateway technology and barcoded sequencing to establish a high-throughput system that is markedly easier to design, highly specific and well-suited for large-scale and multiplexed mutagenesis in a variety of cancer-related genes and pathways.

Understanding cancer heterogeneity by characterizing mutations at large scale

Genetic heterogeneity underlies phenotypic diversity across cancer patient populations¹⁶. The Cancer Genome Atlas (TCGA)¹⁷, the Catalogue of Somatic Mutations in Cancer (COSMIC)¹⁸, and the International Cancer Genome Consortium (ICGC)¹⁹ sequencing projects have catalogued millions of rare and common cancer genomic variants. However, only a small fraction of these variants have been functionally characterized²⁰. Age of disease onset, survival rate, drug response, and clinical pathology are more heterogeneous than originally anticipated in cancer patients^{8–10,20}. A critical challenge then is to identify causal disease variants²¹. In cancer, although germline and somatic variants have been identified in cancer-associated genes²², it is difficult to distinguish driver mutations from the numerous passenger mutations that accumulate during tumorigenesis. Even for a well-established cancer gene such as *TP53*, not all mutations are functionally equivalent²³, and the molecular basis behind disease for most identified mutations remains elusive.

Functional variomics to distinguish various cancer mutations

Genes and gene products interact with each other in signaling networks to carry out cellular and physiological functions^{9,10,20}. Biological networks and cellular systems underlie most genotype-to-phenotype relationships, and perturbations in such networks lead to human disease^{10,24–26}. Different variants of the same gene often give rise to different cancer phenotypes^{27,28}. Gene knockout (such as CRISPR^{29–31}) or knockdown (such as RNAi^{32,33}) technologies for characterization of gene function cannot resolve diverse disease manifestations caused by heterogeneous mutations of the same gene. Genomic site-specific CRISPR knock-in approach³⁴ is promising, but currently not amenable to genome-wide mutation-targeting applications. Therefore, therapeutic and diagnostic strategies urgently require a systems-level understanding of causal and mechanistic links between genetic variation and disease outcomes in context of signaling networks.

Pathogenic mutations in human cancer may exert different effects on cellular networks, ranging from loss of all the interactions of the gene (“quasi-null”), to specific perturbation of one or several but not all molecular interactions (“edgetic”), to retention of all wild-type interactions (“quasi-wildtype”)^{35,36}. Systematic experimental efforts have just emerged to assess the extent to which network perturbations could account for disease phenotypes^{37,38}. Molecular interaction changes mediated by patient-specific mutations have been profiled to investigate at large scale the underlying disease mechanisms^{15,39,40}. By characterizing molecular interactions, the first-version of a functional landscape for human disease mutations was charted¹⁵. For protein-protein interactions only, while common variants from healthy individuals rarely affect interactions, strikingly, ~30% disease mutants exhibit selective perturbations of their wild-type interactions, whereas another ~30% result in loss of all interactions, likely due to defects in protein folding^{15,41}. Together, these results provide evidence for widespread mutation-induced interaction-specific network perturbations across a broad spectrum of diseases^{15,39}.

Here we explain in detail how to use a Gateway technology-integrated, barcoded sequencing-based platform¹⁵ to facilitate massively parallel site-directed mutagenesis at high efficiency and quality (Fig. 1). We further describe functional assays to systematically dissect and stratify large numbers of cancer mutations, and to prioritize driver events. We include considerations for preparing error-free wild-type template sequences, protocols for rapid construction and verification of specific cancer mutations of interest, and finally the use of the resulting mutation Entry clones for functional characterization and stratification in yeast cells and human embryonic kidney (HEK 293) cell lines. The system can similarly be applied to generation and characterization of mutations involved in other human diseases, cell types and organisms, and is amenable to both small-scale and large-scale applications.

Applications and comparison with other site-directed mutagenesis technologies

As with other site-directed mutagenesis methods such as PFunkel, deep mutational scanning and Clone-seq, the emerging Gateway-mediated functional variomics technology¹⁵ can facilitate high-throughput mutagenesis at specific loci of interest in the genome (Fig. 2) and enable easy downstream functionalization via Gateway transfer. In addition to the efficient and high-quality mutagenesis, this platform ensures broad applications due to the Gateway

technology, such as DNA-, RNA-, and protein-based assays as well as molecular interaction assays. Functional variomics offers several potential advantages over previous technologies, including higher efficiency, the ease of customization, the ability to facilitate multiplex mutagenesis and easy functional stratification of various mutations, as described in the following sections:

Higher efficiency.—Although powerful, early approaches are limited in their scope to creating and examining, at most, hundreds of DNA variants. This new functional variomics strategy takes advantage of Gateway technology⁴² that can ease the transfer of mutations to a variety of destination vectors for functional stratification, and thousands of (or more, tens of thousands of) variants can be examined.

Flexible customization.—Previous technologies often have lengthy procedures to generate mutations (such as PFunkel¹¹), or require *de novo* synthesis of relatively long fragments (hundreds of base pairs for Gibson assembly-based methods, such as deep mutational scanning^{12,13}). Functional variomics begins with high-quality sequence-verified ORFeome sequences⁴³ as wild-type template, therefore it is cost-effective and can target 10,000~20,000 genes of interest for mutagenesis.

Barcoded sequencing.—Although deep mutagenesis strategies often use next-gen sequencing for mutation confirmation, most of them are limited by the size of the sequencing reads. For instance, deep mutational scanning currently limits the size of the region that can be mutagenized to a maximum of 300 amino acids¹³. Clone-seq places different mutations of the same gene in one pool for sequencing, and therefore cannot distinguish which mutant clone has unwanted mutations¹⁴. For further confirmation, this technology needs multiple subsequent Sanger sequencing runs to determine unwanted mutations. By contrast, functional variomics uses barcoded sequencing^{15,44} to distinguish different mutations of the same gene, and can easily trace unwanted mutations to particular single colony clones.

High-resolution characterization.—Most site-directed mutagenesis approaches do not have a functional stratification step for characterizing mutations, except for the deep mutational scanning platform¹². However, this method cannot be applied to proteins with complex functions or no known functions. In marked contrast, the functional variomics platform can distinguish a large number of mutations at the fundamental molecular interaction level^{45,46}, and at single base resolution^{36,47}.

Limitations of the functional variomics system

Functional variomics is a robust site-directed mutagenesis pipeline that can efficiently construct a large library (at least thousands) of disease-specific mutations, and experimentally dissect the functional role of each mutation. Since the wild-type templates are from the ORFeome, the mutations in genes that are not available in the ORFeome would be not possible to generate. However, given the broad coverage and ever-expanding nature of human ORFeome^{45,48}, this requirement does not severely limit the capability of functional variomics. Another possible limitation is the technical challenge of cloning mutations in

high GC-content or low complexity regions. This might lead to non-specific off-target events, but optimization of primer design⁴⁹ could alleviate this issue.

Experimental design

High-throughput generation of human cancer mutation Gateway Entry clones.

—Human cancer mutations can be cloned at large scale using an enhanced high-throughput site-directed mutagenesis pipeline^{15,45} (Fig. 2). First, the corresponding full-length wild-type reference genes can be obtained from their Entry clones in human ORFeome v8.1⁴³. These wild-type Entry clones can be transferred into the pDEST-AD vector (AD domain from the yeast transcription factor Gal4) to serve as template for mutagenesis. This is followed by 3 PCR experiments. For a given cancer mutation, this process consists of two “primary PCRs” to generate gene fragments flanking the mutation site, and one “fusion PCR” to obtain the full-length gene harboring the desired mutation. To create mutation Entry clones, all the mutants can be cloned into the Gateway donor vector, pDONR223, by a BP (standing for ‘attB × attP recombination’) reaction followed by bacterial transformation. Independent colonies for each mutant can be isolated for sequence confirmation.

Confirmation of mutations using barcoded next-generation sequencing.—

To confirm expected mutations, mutant clones are PCR amplified from each mutation Entry clone colony (Fig. 3). PCR products can be processed for next-generation sequencing^{15,45}. When multiple sets of mutations are being generated, distinct barcodes can be included for sequence verification (Fig. 3). This makes it possible to combine multiple experiments in a single sequencing run, and separate them computationally afterward. For large-scale experiments, this approach is more cost-effective than traditional Sanger sequencing. Moreover, there is less loss of verified mutations due to failed sequencing, since each mutation PCR product is sequenced multiple times. The clones that have a full-length coverage with only the single desired mutation can be selected and consolidated.

Large-scale functional characterization of mutant clones.—In recent years, large-scale functional variomics platforms have emerged to characterize mutant clones in a high-throughput manner. The high-throughput enhanced yeast two-hybrid (HT-eY2H) system is one of the most powerful functional approaches available to date for this purpose (Fig. 4). HT-eY2H can be used to explore binary direct protein-protein interactions: it is cost effective, and can be applied on a proteome-wide scale. This system is based on the reconstitution of a functional transcription factor through the interaction between two proteins: one fused to the DNA binding domain of the transcription factor (DB-X), and one fused to the activation domain (AD-Y)⁵⁰. Several reporter genes can be used to detect an interaction between DB-X and AD-Y, including selectable, auxotrophic, markers such as *HIS3* (ref. ⁵¹). To accomplish this as stringently and efficiently as possible, an optimized protocol can be implemented to test each candidate interaction in replicate using high-stringency criteria to remove potential false positives^{15,45}. Finally, the identity of the clones exhibiting protein-protein interaction phenotypes can be confirmed using next-generation sequencing⁵².

High-throughput enhanced yeast two-hybrid (HT-eY2H) system.—Once available, the cancer mutant Entry clones and their corresponding wild-type counterparts can be transferred by Gateway LR (standing for ‘attL × attR recombination’) reactions into the HT-eY2H vectors pDEST-DB and pDEST-AD, expressing the yeast Gal4 DNA-binding domain fusion proteins (DB-X), and activation domain fusion proteins (AD-Y). The DB-X and AD-Y plasmids can then be transformed into the haploid yeast Y8930, mating type *MATa*, with the genotype *leu2–3,112 trp1–901 his3 200 ura3–52 gal4 gal80 GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2cyh2^R*, and selected on synthetic complete (SC) agar media without leucine (SC-Leu) to generate Y2H bait strains⁵¹. The prey strains expressing AD-Y can be transformed into the yeast Y8800, mating type *MATa*, with similar genotype as Y8930, and selected on SC media without tryptophan (SC-Trp)⁴⁵.

Protein-protein interaction (PPI) pairwise test.—Individual DB-X and AD-Y yeast strains are inoculated into liquid cultures in SC-Leu and SC-Trp, respectively (Fig. 4). These yeast cells are then mated in liquid YPD media. After mating, the yeast culture is transferred into SC-Leu-Trp liquid media to enrich for diploid cells. The diploid yeast cells are then robotically spotted onto both selective and control agar plates, SC-Leu-Trp-His+3AT and SC-Leu-His+3AT+CHX, respectively. After incubation, PPIs are considered positive for activation of *GAL1::HIS3* (i.e., growth to overcome 3AT inhibition) in a CHX-sensitive manner. PPI comparisons of mutant clones with their wild-type controls give rise to mutation-mediated PPI perturbations. Note, the counter-selectable CYH2 marker is important for identification and elimination of DB-X auto-activators⁵³ that could activate the *GAL1::HIS3* reporter gene expression in the absence of an AD-Y plasmid.

Validation of interaction network perturbations by *Gaussia princeps* Protein-fragment Complementation Assay (GPCA) in mammalian cells.—Mutation-induced interaction perturbations can be validated with high-throughput Gateway-compatible *Gaussia princeps* protein-fragment complementation assay (GPCA) (Fig. 5). GPCA detects protein-protein interactions by fragment complementation of *Gaussia princeps* luciferase⁵⁴. pSPICA-C1 and pSPICA-N2 expression vectors each contain one of the two fragments of the humanized *Gaussia princeps* luciferase (hGLuc F[1] and hGLuc F[2], respectively)^{54,55}. The tested protein pairs can be transferred into the pSPICA-C1 and pSPICA-N2 vectors by LR clonase-mediated Gateway reaction. As a result, the hGLuc F[1] luciferase fragment is linked to the C-terminus of a tested protein, and the complementary hGLuc F[2] fragment is linked to the N-terminus of a protein partner. Both mammalian expression vectors carry the same human cytomegalovirus (CMV) promoter and are maintained as high copy number with the human virus SV40 replication origin.

Interaction scoring and data analysis.—Protein pairs tested are defined as valid only if both partners are successfully cloned into expression vectors, and both are transformed or transfected successfully into cells. The average interaction score of independent experimental repeats is considered the final interaction score for each protein pair. The protein pairs are tested together with the positive and negative control pairs on the same experimental plates. A quantitative readout can be used to titrate the “threshold” signal to an acceptable range. The threshold is defined so that any pair scoring above that threshold is

considered “positive” and otherwise considered “negative”. The threshold is set such that the false positive recovery rate is as minimal as 1% to 2.5%.

MATERIALS

REAGENTS

Antibiotics

- Spectinomycin (Sigma-Aldrich, cat. no. S4014)
- Carbenicillin (Sigma-Aldrich, cat. no. C1389)

Reagents and media

- Agar (Sigma-Aldrich, cat. no. A1296)
- D-(+)-glucose (Sigma-Aldrich, cat. no. G8270)
- Glycine (Sigma-Aldrich, cat. no. G7126)
- Ethanol, absolute (200 proof, Fisher Scientific, cat. no. BP28184) **CAUTION**
Ethanol is highly flammable. Handle it in a fume hood. Ensure adequate ventilation.
- L-alanine (Sigma-Aldrich, cat. no. A7627)
- L-arginine (Sigma-Aldrich, cat. no. A5006)
- L-asparagine (Sigma-Aldrich, cat. no. A8381)
- L-aspartic acid (Sigma-Aldrich, cat. no. A9256)
- L-cysteine (Sigma-Aldrich, cat. no. C7352)
- L-glutamic acid (Sigma-Aldrich, cat. no. G5889)
- L-glutamine (Sigma-Aldrich, cat. no. G3126)
- L-histidine (Sigma-Aldrich, cat. no. H8000)
- L-isoleucine (Sigma-Aldrich, cat. no. I2752)
- L-leucine (Sigma-Aldrich, cat. no. L8000)
- L-lysine (Sigma-Aldrich, cat. no. L5626)
- L-methionine (Sigma-Aldrich, cat. no. M9625)
- L-phenylalanine (Sigma-Aldrich, cat. no. P2126)
- L-proline (Sigma-Aldrich, cat. no. P0380)
- L-serine (Sigma-Aldrich, cat. no. S4500)
- L-threonine (Sigma-Aldrich, cat. no. T8625)
- L-tryptophan (Sigma-Aldrich, cat. no. T0254)
- L-tyrosine (Sigma-Aldrich, cat. no. T3754)

- L-valine (Sigma-Aldrich, cat. no. V0500)
- Ammonium sulfate (Sigma-Aldrich, cat. no. A4418)
- Yeast peptone (Sigma-Aldrich, cat. no. 39396)
- Yeast extract (EMD Chemicals, cat. no. 1.03753)
- Yeast Nitrogen Base (without amino acids, without ammonium sulfate) (Sigma-Aldrich, cat. no. 51483)
- Salmon sperm DNA, 10 mg/ml (Sigma-Aldrich, cat. no. D7656)
- Cycloheximide (Sigma-Aldrich, cat. no. C7698)
- Adenine (Sigma-Aldrich, cat. no. A3159)
- Uracil (Sigma-Aldrich, cat. no. U0750)
- SOC medium (New England BioLabs, cat. no. B9020S)
- 3-AT (3-Amino-1,2,4-triazole) (Sigma-Aldrich, cat. no. A8056). **CAUTION** 3-AT is a harmful chemical. Avoid dust formation. Avoid breathing dust. Ensure adequate ventilation.

Mammalian cell culture

- Polyethylenimine (PEI) “Max” (Polysciences Inc., cat. no. 24765)
- *Renilla* Luciferase Assay system (Promega, cat. no. E2820)
- DMEM-GlutaMAX™ (Life Technologies, cat. no. 10569–010)
- FBS (Life Technologies, cat. no. 10438–034)
- Penicillin-streptomycin (Life Technologies, cat. no. 15140–163)
- Human embryonic kidney (HEK) 293T cells **CAUTION** The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

Biochemical enzymes

- KOD hot start polymerase (EMD Chemicals, cat. no. 71086) **CRITICAL** To minimize error in amplifying DNAs during mutagenesis, it is important to use a high-fidelity polymerase.
- HiFi Platinum Taq polymerase (Life Technologies) **CRITICAL** To minimize error in amplifying DNAs during mutagenesis, it is important to use a high-fidelity polymerase.
- Restriction enzymes (NEB)

Gateway kits

- Gateway® BP Clonase® Enzyme mix (Life Technologies, cat. no. 11789–021) **CRITICAL** Always use these enzymes right before experiments, and make

aliquots to avoid multiple freeze-thaw cycles. For long-term storage of Gateway clonase enzymes, store at -80°C .

- Gateway® LR Clonase® Enzyme mix (Life Technologies, cat. no. 11791–043) **CRITICAL** Always use these enzymes right before experiments, and make aliquots to avoid multiple freeze-thaw cycles. For long-term storage of Gateway clonase enzymes, store at -80°C .

Chemicals for buffers

- Glycerol (Sigma-Aldrich, cat. no. G5516)
- NaH_2PO_4 (Sigma-Aldrich, cat. no. S3264)
- Na_2HPO_4 (Sigma-Aldrich, cat. no. S3139)
- Tris (Trizma Base) (Sigma-Aldrich, cat. no. T1503)
- EDTA (Sigma-Aldrich, cat. no. E6758)
- Lithium Acetate (LiAc) (Sigma-Aldrich, cat. no. L6883)
- PEG 3350 (Sigma-Aldrich, cat. no. P3640)

Primers

- Custom DNA primers, 50 μM in a 96-well format, no PAGE purification required (See Table 1; Sigma-Aldrich)

EQUIPMENT

- Round-bottom 96-well plates (Corning, cat. no. 3788)
- 50 ml sterile reagent troughs (Corning, cat. no. 4871)
- Thin-walled 96-well PCR microtiter plates (Bio-Rad, cat. no. HSP-9601)
- 15 cm sterile petri dishes (Fisher Scientific, cat. no. 08-757-14)
- Aluminum foil microplate seal (Bio-Rad, cat. no. MSF-1001)
- 50 ml polypropylene tubes (BD Falcon, cat. no. 352070)
- Tissue culture 96 wells flat bottom plate (BD Falcon, cat. no. 353075)
- PIPETMAN P10, 1 to 10 μl (Gibson, cat. no. F144802)
- PIPETMAN P200, 50 to 200 μl (Gibson, cat. no. F123601)
- Corning™ 4084 P10 multichannel pipettes 1 - 10 μl (Thermofisher scientific, cat. no. 07-764-710)
- Corning™ 4085 P50 multichannel pipettes 5 - 50 μl (Thermofisher scientific, cat. no. 07-764-711)
- Corning™ 4085 P200 multichannel pipettes 20 - 200 μl (Thermofisher scientific, cat. no. 07-764-712)

- Corning® Stripettor™ Ultra Pipet Controller (Thermofisher scientific, cat. no. 4099)
- Thermocycler capable of handling 96-well plates (Bio-Rad, cat. no. T100™)
- Temperature controlled incubator (30°C) (Thermofisher scientific, cat. no. 15-103-0515)
- Temperature controlled water bath (42°C) (Thermofisher scientific, cat. no. FSGPD10)
- 1% E-Gel® 96 Agarose Gels (Thermofisher scientific, cat. no. G700801)
- E-gel Mother E-Base capable of running 96 samples (Thermofisher scientific, cat. no. EBM03)
- Temperature controlled shaker for 250 ml flasks and 50 ml tubes (MaxQ™ 6000 Shaker, Thermofisher scientific, cat. no. SHKE6000)
- Benchtop Centrifuge suitable for microplates and 50 ml tubes (Sorvall™ Legend™ XT/XF centrifuge, Thermofisher scientific, cat. no. 75-004-505)
- Plate shaker (Mixmate, Eppendorf, cat. no. 022674200)
- GENESYS™ 10S UV-Vis Spectrophotometer (Thermofisher scientific, cat. no. 840-208100)
- Vortexer (Thermofisher scientific, cat. no. 88880017TS)
- Centro XS3 Luminescence Microplate Reader (Berthold Technologies, cat. no. 46970)
- 1300 Series A2 Class II, Type A2 Bio Safety Cabinets (Thermofisher scientific, cat. no. 13-261-221)
- Optional: liquid handling robot with 96-well multichannel head and individual well cherry-picking capability (Tecan, cat. no. Fluent 780)
- Optional: BioRobot Universal System Robot with 96-well DNA miniprep capability (Qiagen, cat. no. 9001094)

REAGENT SETUP

2xYT—Prepare 900 ml of doubly distilled H₂O (ddH₂O). Add 16 g of Bacto Tryptone, 10 g of Bacto Yeast Extract, and 5 g of NaCl. Adjust pH to 7.0 with 5N NaOH. Bring the final volume to 1 L. 2xYT medium can be sterilized by autoclaving. This solution can be stored at 4 °C for up to 6 months.

SOC medium—0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose. Add glucose after autoclaving the solution with the remaining ingredients, and let cool down. Sterilize the final solution by passing it through a 0.2 µm filter. Medium can be stored at –20 °C for months.

40% glucose (w/v)—This solution is made by first bringing 500 ml of doubly distilled H₂O (ddH₂O) to a boil, then stirring in 400 g of glucose, and adjusting the final volume to 1 L. Glucose solution can be sterilized by autoclaving. Remove glucose promptly from the autoclave when pressure allows to prevent caramelization. This solution can be stored at 4 °C for up to 6 months.

Sodium phosphate buffer—Prepare 190 ml of 0.1 M NaH₂PO₄ and 810 ml of 0.1 M Na₂HPO₄. Mixing these should result in a 0.1 M phosphate buffer with a pH of 7.4. This solution can be stored at room temperature (20~25 °C) for 1 year.

Yeast lysis buffer—Dissolve 2.5 mg/ml of zymolase (20T) in sodium phosphate buffer. Prepare 10 µl per lysis reaction. Keep on ice before use. This lysis buffer can be stored at -20 °C for months.

Synthetic complete (SC) dropout amino acid mixture—SC liquid media and plates are based on an amino acid mixture lacking leucine, tryptophan, and histidine. When necessary, leucine, tryptophan, or histidine, as well as the additives cycloheximide and 3-AT are added after autoclaving. The amino acid mixture can be generated by mixing equal amounts of the following components: alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, and valine. This mixture can be stored at room temperature for 1 year.

SC stock solutions—100 mM histidine in H₂O; 100 mM leucine in H₂O; 40 mM tryptophan in H₂O (store covered with aluminum foil at 4°C); adenine (12 g/L); 20 mM uracil in H₂O; 10 mg/ml cycloheximide in ethanol (10,000× stock). These solutions can be stored at 4 °C for up to 6 months.

SC media and plates—For 1 L of SC medium, dissolve in 950 ml distilled water 1.5 g of SC dropout amino acid mixture, 2 g of yeast nitrogen base (without amino acids), and 5 g of ammonium sulfate. Adjust the pH to 5.9 with 5N NaOH, then add 20 g agar if making plates. After autoclaving, add 50 ml of 40% glucose, 15ml of 12 g/L adenine, and 8 ml of 20 mM uracil. If necessary, add the following additionally required ingredients: 8 ml each of stock solution of amino acid (leucine, tryptophan, or histidine). Add the required amount of 3-AT (84 mg/L for 1 mM) and cycloheximide (100 µl of stock solution for a final concentration of 1 mg/L). 1 L suffices for ten 15-cm Petri dishes. This medium can be stored at 4 °C for up to 6 months.

YPD media and plates—For 1 L, dissolve 20 g of bacto peptone and 10 g of yeast extract in 950 ml of distilled water. Add 20 g of agar if making plates. Autoclave, then add 50 ml of 40% glucose and 15 ml of adenine (12 g/L), and mix. This medium can be stored at 4 °C for up to 6 months.

10× TE buffer: 100 mM Tris, 10 mM EDTA, pH 8.0. This solution can be stored at room temperature for 1 year.

M solution of LiAc.: Sterilize by autoclaving. This solution can be stored at room temperature for 1 year.

50% w/v polyethylene glycol (PEG).: Dissolve 125 g of PEG in warm ddH₂O and adjust the final volume to 250 ml. Filter sterilize the 50% PEG solution (this will take 5-10 min).

Salmon sperm DNA.: Just before transforming yeast cells, thaw the 10 mg/ml salmon sperm DNA, then boil for 5 minutes to denature the DNA, and place on ice bath to prevent re-annealing.

TE/LiAc/PEG solution.: Prepare by mixing 1/10 of final volume of 10× TE with 1/10 final volume of 1 M LiAc, and filling up to desired volume with 50% PEG. Freshly prepare the reagents for each experiment.

PROCEDURE

Large-scale computational design of cancer mutation cloning primers •

TIMING 1 d—1 For all mutation clones that need to be created, design mutation-specific oligonucleotide primers (37-45 nucleotides long): such that the first 18-22 nucleotides correspond to the 5' sequence flanking the mutation, then the mutation site, and the remaining 18-22 nucleotides match the 3' flanking sequence. Corresponding sets of forward and reverse primer plates should be ordered (*e.g.*, for a particular mutation, the forward primer is in plate 1 well A1, and the reverse primer in plate 2 well A1). Wild-type primers are not necessary.

CRITICAL STEP A decision often has to be made whether to include a stop codon in the primer, or have the vector provide the stop codon. The human ORFeome entry clones do not contain stop codons. This enables easy transfer into vectors harboring a C-terminal tag. However, if the destination vectors do not have a stop codon, then it is better to include a stop codon in the reverse primer.

2 Dilute forward and reverse mutation-specific primers from **step 1** in 96-well plates (*e.g.*, PCR plates) to a final concentration of 1 μ M each, in a volume of 100 μ l double distilled water (ddH₂O).

Preparation of mutagenesis PCR templates • TIMING 3 d—3 The human ORFeome contains a comprehensive collection of full-length wild-type entry clones. To prepare templates for PCR in our mutagenesis pipeline, we need to transfer wild-type entry clones into the destination vector, pDEST-AD. Perform enzymatic digestion of pDEST-AD vector as follows:

Component	Volume (μ l)	Final concentration
pDEST-AD (1 μ g/ μ l)	50	100 ng/ μ l
NEB CutSmart buffer (10×)	50	1×
SmaI enzyme	5	100 units

Component	Volume (μl)	Final concentration
Double distilled H ₂ O (ddH ₂ O)	395	

4 Pipet well to mix, and incubate at 25 °C for 12-24 h.

5 Heat inactivate the enzyme by incubation at 65 °C for 20 min. Run 1 μl of digested pDEST-AD and 1 μl of undigested pDEST-AD as a negative control on a 0.8% agarose gel. Run the gel at 120 volts for 30 min, stain with EtBr, and visualize under a UV light. Observe a shift in DNA size to confirm complete digestion.

6 Prepare a master mix for ~100 Gateway LR reactions as follows to transfer genes from entry clones into pDEST-AD:

Component per 96 wells	Volume (μl)	Final concentration
Digested pDEST-AD from Step 5 (100 ng/μl)	110	20 ng/μl
LR clonase buffer (5×)	110	1×
1× TE (pH=8.0)	55	0.1×
LR clonase enzyme mix	55	

7 Pipet well to mix, and distribute 3 μl into each well of a 96-well plate. Add 2 μl of wild-type entry clone plasmids (50-100 ng/μl) per well. After a brief spin down, incubate at 25 °C for 4-12 h.

8 Thaw 1 ml of DH5α or DH10B bacterial competent cells on ice. Add 10 μl of cells into each well of a 96-well plate, and then add 1 μl of Gateway LR reaction products from Step 7 into each well of competent cells. Seal the plate with heat sealer or aluminum foil, and incubate on ice for 30 min.

9 Heat shock at 42 °C in a standard thermocycler for 1 min, and incubate on ice for 2 min. **CRITICAL STEP:** Note, this step is not necessary if using Z-competent cells (Zymo Research, T3020).

10 Add 40 μl of pre-warmed (37 °C) SOC media into each well. Seal the plate with heat sealer or aluminum foil to avoid contamination. Then incubate at 37 °C for 1 h.

11 Transfer each well of the transformed bacterial cultures into a well of a 96-well deep-well plate containing 1.1 ml of 2xYT media with 100 μg/ml of carbenicillin (or ampicillin). Incubate at 37 °C with orbital shaking at 200 rpm speed for 24 h.

12 Verify the identity of LR reaction products by bacterial PCR. Set up the reaction as follows:

Component	Volume (μl)	Concentration in PCR reaction
ddH ₂ O	4.9	-
KOD buffer (10×)	1	1×
dNTP mix (2 mM each)	0.6	0.12 mM each
MgSO ₄ (25 mM)	0.6	1.5 mM
AD primer (1 μM, diluted from 200 μM stock)	1	0.1 μM
Term primer (1 μM, diluted from 200 μM stock)	1	0.1 μM
KOD hot-start polymerase (2.5 U/μl)	0.1	0.025 U/μl

13 Using a reagent trough and a multichannel pipette, distribute 9.2 μl of the PCR master mix from Step **12** per well into thin-walled 96-well PCR microtiter plates.

14 Using a multichannel pipette, add 0.8 μl of bacterial culture from Step **11** to each well as DNA template, for a final reaction volume of 10 μl. Seal the PCR plate using a heat sealer or aluminum foil.

CRITICAL STEP Wells G12 and H12 are used as negative control (water as template) and positive control (10 ng of empty pDEST-AD), respectively.

15 Carry out the PCR as follows. Set the lid temperature to 104°C to prevent condensation of the sample on the plate seal.

Step	Type	Temperature	Duration
1	Denaturation	95 °C	5 min.
2	Denaturation	94 °C	30 sec.
3	Annealing	57 °C	30 sec.
4	Elongation	72 °C	1 min. per kb product
5	Go to step 2	-	33 ×
6	Final elongation	72 °C	8 min.
7	Hold	10 °C	forever

TROUBLESHOOTING

16 If the PCR is successful as determined by running a 0.8% agarose gel to observe the expected DNA sizes, prepare glycerol stock plates by combining 80 μl of sterile 40% w/v glycerol with 80 μl of bacterial culture in 96-well round bottom plates. Mix well. Store plates at –80°C.

PAUSE POINT Glycerol stocks can be stored at –80°C indefinitely.

17 For the remaining bacterial culture in 96-well deep-well plates, perform miniprep plasmid isolation using a QIAgen 96-well plasmid extraction kit, using a bench vacuum setting or a QIAgen liquid handling robot, according to manufacturer's instructions.

Measure DNA concentration by a high-throughput NanoDrop 8000 spectrophotometer. A concentration over 30 ng/μl is expected to continue to the next step.

High-throughput generation of DNA fragments flanking the mutation site •

TIMING 4 h—18 Dilute the wild-type pDEST-AD plasmids with ddH₂O at 1:50 to use as PCR template.

19 To amplify the 5' fragment, prepare a PCR master mix with all components except reverse primer and template as detailed below. Generate enough mixture for all reactions, plus an additional 10% to allow for loss during pipetting steps.

Component	Volume (μl)	Concentration in PCR reaction
ddH ₂ O	2.8	-
KOD buffer (10×)	1	1×
dNTP mix (2 mM each)	0.5	0.1 mM each
MgSO ₄ (25 mM)	0.6	1.5 mM
AD-Tag1 primer (1 μM, diluted from 200 μM stock)	1	0.1 μM
KOD hot-start polymerase (2.5 U/μl)	0.1	0.025 U/μl

CRITICAL STEP The use of a hot-start polymerase allows one to do all the PCR set up at room temperature rather than on ice, which is cumbersome when dealing with multiple plates of PCR reactions. The use of a proofreading enzyme such as KOD ensures high-fidelity amplification.

20 To amplify the 3' fragment, prepare a PCR master mix with all components except forward primer and template as detailed below. Generate enough mixture for all reactions, plus an additional 10% to allow for loss during pipetting steps.

Component	Volume (μl)	Concentration in PCR reaction
ddH ₂ O	2.8	-
KOD buffer (10×)	1	1×
dNTP mix (2 mM each)	0.5	0.1 mM each
MgSO ₄ (25 mM)	0.6	1.5 mM
Term-Tag2 primer (1 μM, diluted from 200 μM stock)	1	0.1 μM
KOD hot-start polymerase (2.5 U/μl)	0.1	0.025 U/μl

21 Using a reagent trough and a multichannel pipette, distribute 6 μl of the PCR master mixes per well into thin-walled 96-well PCR microtiter plates.

22 Using a multichannel pipette or a liquid handling robot, add 2 μl of diluted wild-type pAD-gene template and 2 μl of mutation reverse primer from **step 2** to the distributed 5' fragment PCR master mix or 2 μl of mutation forward primer from **step 2** to the distributed 3' fragment PCR master mix, for a final reaction volume of 10 μl and a final primer

concentration of 0.1 μM each primer. Seal the PCR plate using a heat sealer or aluminum foil.

23 Carry out the PCRs as follows. Set the lid temperature to 104°C to prevent condensation of the sample on the plate seal.

Step	Type	Temperature	Duration
1	Denaturation	95 °C	5 min.
2	Denaturation	94 °C	30 sec.
3	Annealing	57 °C	30 sec.
4	Elongation	72 °C	1 min. per kb product
5	Go to step 2	-	33 ×
6	Final elongation	72 °C	8 min.
7	Hold	10 °C	forever

TROUBLESHOOTING

24 Verify amplification of PCR products by displaying 2 μl of product on a 0.8% agarose gel. Run the gel at 120 volts for 30 min, stain with EtBr, and visualize under a UV light.

PAUSE POINT PCR products can be stored for several days at 4°C or for several weeks at −20°C.

High-throughput mutagenesis of cancer variants by fusion PCR • TIMING 4 h

—**25** Add 2–3 μl each of the 5' and 3' fragment PCR products generated in **step 23** respectively into 100 μl of ddH₂O to serve as fusion PCR templates.

CRITICAL STEP These templates should be prepared right before performing fusion PCR.

26 Fusion PCR: Generate a PCR master mix for the fusion PCR containing all components except template, as described below. Generate enough mixture for all reactions, plus an additional 10% to allow for loss during pipetting steps.

Component	Volume (μl)	Concentration in PCR reaction
ddH ₂ O	10	-
KOD buffer (10×)	2	1×
dNTP mix (2 mM each)	2	0.1 mM each
MgSO ₄ (25 mM)	0.8	1.0 mM
Tag1 primer (1 μM , diluted from 200 μM stock)	1.5	0.075 μM
Tag2 primer (1 μM , diluted from 200 μM stock)	1.5	0.075 μM
KOD hot-start polymerase (2.5 U/ μl)	0.2	0.025 U/ μl

CRITICAL STEP The use of a hot-start polymerase allows one to do all the PCR set up at room temperature rather than on ice, which is cumbersome when dealing with multiple

plates of PCR reactions. The use of a proofreading enzyme such as KOD ensures high-fidelity amplification.

27 Using a reagent trough and a multichannel pipette, distribute 18 µl of the PCR master mix per well into thin-walled 96-well PCR microtiter plates.

28 Using a multichannel pipette or a liquid handling robot, add 2 µl of the diluted fusion PCR template generated in **step 25** to the distributed PCR master mix, for a final reaction volume of 20 µl. Seal the PCR plate using a heat sealer or aluminum foil.

29 Carry out the PCR as follows. Set the lid temperature to 104°C to prevent condensation of the sample on plate seal.

Step	Type	Temperature	Duration
1	Denaturation	95°C	5 min.
2	Denaturation	94°C	30 sec.
3	Annealing	58°C	30 sec.
4	Elongation	72°C	1 min. per kb product
5	Go to step 2	-	30 ×
6	Final elongation	72°C	8 min.
7	Hold	10°C	forever

TROUBLESHOOTING

30 Verify amplification of fusion PCR products by displaying 2 µl of product on a 0.8% agarose gel. Run the gel at 120 volts for 30 min, stain with EtBr, and visualize under a UV light.

PAUSE POINT PCR products can be stored for several days at 4°C or for several weeks at −20°C.

TROUBLESHOOTING

Large-scale construction of cancer mutation entry clones • TIMING 2-3 d—31

Perform Gateway BP reactions to transfer mutation fusion PCR products into the Gateway Donor vector pDONR223. Set up 100 BP reactions as follows:

Component per 96 wells	Volume (µl)	Final concentration
pDONR223 (100 ng/µl)	110	20 ng/µl
BP clonase buffer (5×)	110	1×
1× TE (pH=8.0)	55	0.1×
BP clonase enzyme mix	55	

32 Repeat Step 7, using 2 µl of mutation fusion PCR products per well from Step 30. Then repeat Step 8 using 1 µl of Gateway BP reaction products from step 31, followed by Steps 9 & 10.

33 To get single colonies, spot 5 µl of the transformed bacterial culture onto a 15-cm LB agar plate containing 100 µg/ml of spectinomycin. Incubate at 37 °C with orbital shaking at 200 rpm speed for 24 h.

CRITICAL STEP To make sure to get single colonies rather than a patch, it is better to also perform serial dilutions with H₂O of the transformed bacterial culture (such as 1:10), and then spot 5 µl onto a 15-cm LB+ spectinomycin agar plate.

34 Verify the identity of BP reaction products by bacterial PCR by repeating Steps **12-16**, using M13F and M13R primers in Step **12**, and 10 ng of empty pDONR223 as positive control in Step **14**.

TROUBLESHOOTING

Verification of mutation entry clones by barcoded next-generation sequencing

• **TIMING 5-6 d hands-on, 2-3 weeks expansion—CRITICAL** All mutation entry clones need to be verified, usually by barcoded pooling and next-generation sequencing technologies. Note that if only a few mutant clones need to be confirmed, traditional Sanger sequencing with walking primers⁵⁶ may be more cost effective, and can be used to sequence the PCR products generated in **step 34**.

35 Make pools of ~1,000 PCR products. This can be done by transferring all PCR products in each well of 96-well plates into a common reservoir, and mixing well by gentle shaking or pipetting. To ensure precise subsequent sequencing analysis, please ensure that only one mutation from each gene is present in each pool. This facilitates the mutant clone sequence assembly downstream of next-generation sequencing.

36 Purify the pooled PCR products using MinElute PCR Purification Kit (Qiagen), following the manufacturer's instructions, and quantify DNA content by a NanoDrop spectrophotometer.

37 Multiplex up to ~100 pools by ligation to unique barcode adaptors, according to manufacturer's instructions. Adaptors can be chemically synthesized, or ordered directly from a sequencing company. Ideal adaptors should be short (10-75 bp), and not exhibit any homology with the human genome. ~100 pools (each containing >1,000 mutations) can be generated by using 96 unique barcodes (a combination of 12 row index adaptors and 8 column index adaptors).

38 Mutation PCR amplicon pools with adaptors linked were further combined together into one sample per library, and quantified by a NanoDrop spectrophotometer.

39 Subject the PCR library for mutation verification by next-generation sequencing technologies (454-FLX, Illumina Solexa, etc), using sequencing kit provided by the

sequencing company. This step can be done by paired-end or single-end sequencing strategy, and following manufacturer's instructions.

40 Assemble reads from next-generation sequencing runs and align them to the reference sequences (wild-type genes). Millions of reads can be obtained per library using HiSeq 2500 or NextSeq 500. After low-quality bases have been trimmed, reads can be aligned to the reference ORF sequences of interest, using *de novo* genome assembly programs, such as DNASTAR.

41 Sequence analysis and assessment: After sequence alignment, mutant clones that have a full-length coverage with only the single desired mutation are considered as confirmed mutants, which can be used for functional studies as described below.

Gateway transfer of cancer mutations into enhanced yeast two-hybrid expression vectors • TIMING 2-3 d—42

Repeat Steps **6** and **7** using 2 µl of mutant entry clones (30-100 ng/µl) from Step **41** per well to perform Gateway LR reactions into pDEST-DB and pDEST-AD vectors, deriving DB-X and AD-Y mutant clones, respectively, followed by repeating Steps **8**, **9**, **10** & **11**.

43 Verify the identity of LR reaction products by bacterial PCR by repeating Steps **12-16**, using DB and Term primers instead for pDEST-DB samples in Step **12**, and 10 ng of empty pDEST-DB as positive control in Step **14**. Then repeat Step **17** to extract plasmid DNA.

TROUBLESHOOTING

Transforming yeast cells with DB-X or AD-Y mutant clones • TIMING 4 d—44

Separately inoculate yeast strains Y8800 (*MATa*) and Y8930 (*MATa*) on YPD agar plates, and grow at 30°C for 24 hrs.

45 Using an inoculation loop or a sterile toothpick, inoculate 20 ml of YPD in a 50 ml polypropylene tube or a 100 ml Erlenmeyer flask with the Y8800 and Y8930 strains grown on YPD agar plates (a separate tube or flask for each strain). Use 5–10 colonies per tube for the inoculation. Grow overnight at 30°C in a shaking incubator at 200 rpm (the final optical density at 600 nm will be 3–5).

46 Dilute each yeast strain in YPD to an optical density at 600 nm of 0.3. Use 100 ml of YPD per 96 transformations, and an Erlenmeyer flask with a volume at least four times the culture volume. Grow the diluted yeast cells at 30°C in a shaking incubator at 200 rpm until the optical density at 600 nm reaches 0.8–1.0 (4–6 hrs).

47 Just before the yeast cells reach the desired optical density at 600 nm, prepare the transformation solutions (see Reagent setup). **CRITICAL STEP** All further transformation steps are done at room temperature unless otherwise indicated.

48 Prepare 4.45 ml of yeast transformation buffer mixtures per 96 transformations as described below and mix vigorously by vortexing.

Transformation buffer reagent	Volume per 96 transformations
1M LiAc	500 μ l
10 $\times$ TE	500 μ l
60% polyethylene glycol (PEG)	3300 μ l
Boiled salmon sperm DNA (10 mg/ml)	150 μ l

49 Harvest yeast cells by centrifugation at $1000 \times g$ for 5 min. Discard supernatant and resuspend cell pellet in 10 ml of sterile H₂O per 100 ml of original culture by vigorous shaking.

50 Centrifuge yeast cells at $1000 \times g$ for 5 min. Discard supernatant and resuspend cell pellet in 4 ml transformation buffer mixture from Step **48** per 100 ml of original culture by vigorous shaking.

51 Using a sterile reagent trough and a multichannel pipette, distribute 40 μ l of the yeast transformation mixture per well into round-bottom 96-well plates.

52 Add 5 μ l of AD-Y or DB-X DNA (250 ng – 1 μ g) from Step **43** into each yeast transformation mixture - use Y8930 (*MAT α*) for all DB-X transformations, and Y8800 (*MAT a*) for all AD-Y transformations. As a negative control, do not add DNA clones to two wells per plate. Mix well by pipetting up and down. Seal plate with a heat sealer or aluminum foil.

53 Apply a heat shock by incubating for 30-45 minutes at 42°C in a water bath.

54 Spot 6 μ l of resuspended yeast cells onto the appropriate selective media plate (SC -Trp for AD-Y and SC -Leu for DB-X). Use of a liquid handling robot is highly recommended to obtain regular spacing of colonies. If using a multichannel pipette, print out a 96-well grid and place the plate on top of this to guide the pipetting.

55 Incubate plates for 48-72 hours at 30°C.

56 Check for efficient transformation. Transformed yeast cells should form a solid patch, while negative controls should be entirely empty.

TROUBLESHOOTING

PAUSE POINT Yeast plates can be kept at 4°C for up to a month, though it is better to continue to the end of Step **59** before pausing.

57 Using sterile toothpicks or sterile pipette tips, pick transformed yeast cells from the selective plates and transfer to individual wells of round bottom 96-well plates containing 110 μ l of liquid selective medium (SC -Trp for AD-Y and SC -Leu for DB-X).

58 Incubate plates for 24 hours at 30°C on a plate shaker.

59 Prepare one set of glycerol stock plates by combining 80 µl of sterile 40% w/v glycerol with 80 µl of yeast culture in 96-well round bottom plates. Mix well. Store plates at –80°C.

PAUSE POINT Glycerol stocks can be stored at –80°C indefinitely.

Characterizing protein-protein interaction profiles of mutant clones • TIMING

5-6 d—60 In the following steps, matching sets of 96-well plates containing haploid DB-X and AD-Y yeast strains are generated and mated together to create the diploid yeast strains expressing protein pairs to be tested. Develop a pairing scheme to distribute the haploid DB-X and AD-Y yeast strains from **Step 56** into 96-well plates in the order decided on before starting the protocol, including the location of positive and negative control pairs.

CRITICAL STEP In each assay plate, include at least two pairs known to interact as positive controls, and two random pairs as negative controls. To avoid any bias, randomize the list of interactions to be tested. This way, the phenotypes can be scored objectively, without knowledge of which locations should test positive or negative.

61 Fill 96-well round-bottom microtiter plates with 110 µl of selective medium (SC–Leu for Y8930 containing DB-X, and SC–Trp for Y8800 containing AD-Y).

62 Using the glycerol stocks prepared in **Step 59** and existing glycerol stocks of positive and negative control yeast strains, inoculate the media plates with 5 µl of yeast cells per well. For accuracy, use of a cherry-picking capable liquid handling robot is highly recommended.

63 Incubate plates for 24 hours at 30°C in a shaking incubator.

64 Using a multichannel pipette or liquid handling robot, mix equal volumes (5 µl) of corresponding sets of DB-X and AD-Y yeast plates in a 96-well round-bottom plate containing 110 µl of YPD medium. Use of a liquid handling robot is recommended to obtain a regular spacing of colonies. If using a multichannel pipette, print out a 96-well grid and place the plate on top of this to guide the pipetting. The remaining haploid yeast cells can be stored as glycerol stocks at –80°C by adding an equal volume of 40% sterile glycerol.

65 Incubate the plates for 12–24 hours at 30°C.

66 Inoculate yeast cells from liquid YPD culture into 110 µl of SC-Leu-Trp liquid media to select for diploid strains from successful yeast mating.

67 Incubate the plates for 24 hours at 30°C.

68 Spot 5 µl of mated yeast cells onto the following assay plates. For each plate that assays activation of a reporter gene, there is a corresponding plate that detects DB-X that activates the reporter in the absence of an AD-Y plasmid (Fig. 4).

Plate type	Purpose
SC –Leu –Trp –His +1 mM 3AT	Assays <i>HIS3</i> reporter activity (stringent condition)

Plate type	Purpose
SC –Leu –His +1 mM 3AT+ 1 mg/l CHX	Corresponding autoactivation detection

69 Onto each plate, spot a series of Y2H controls with interactions yielding known phenotypes (Fig. 4). These controls are used to determine that the assay plates were properly generated and that the phenotypes are scored at the appropriate time.

70 Incubate the plates for 48–72 hours at 30°C.

71 Check that the Y2H controls display the expected phenotypes (Fig. 4) on the assay plates. Then score the growth of the yeast patches on the two types of assay plates. When scoring, examine the behavior of the entire yeast patch. If a high number of colonies arise from a patch on an autoactivation assay plate, score the patch as an autoactivator. Or else, if on the autoactivation assay plate single colonies grow, while on the reporter assay plate it is a full patch growth. These should be scored as positive (not autoactivator).

TROUBLESHOOTING

72 Collate the replicate scores for each protein pair into a final score. A protein pair scores positive for interaction when it shows growth on both repeats on the reporter activity plate, but not on the corresponding autoactivation assay plate.

73 Using a 200 µl pipette tips, from one set of assay plates pick yeast patches that score positive into 96-well PCR microtiter plates containing 110 µl of SC-Leu-Trp-His+3AT liquid media. Use of a liquid handling robot is highly recommended to obtain a regular spacing of colonies.

74 Incubate the liquid media plates for 24 hours at 30°C.

PAUSE POINT Although fresh yeast plates work best for lysis and PCR, the picked yeast cells can be stored at –80°C for years by mixing 80 µl of 40% glycerol with 80 µl of cells.

Lysis of yeast cells for PCR and sequence verification of positive interaction pairs • TIMING 2h—75 Distribute 10 µl of lysis buffer per well into 96-well PCR microtiter plates.

76 Using 10 µl pipette tips, take out 1 µl of yeast cells from the well bottom of the liquid SC-Trp-Leu-His+3AT media plate from Step **74**, and resuspend yeast cells in the lysis buffer. With practice, this can be done with a multichannel pipette, 12 wells at-a-time. Seal the PCR plate using a heat sealer or aluminum foil.

77 Run the lysis reaction on a thermocycler as follows. Set the lid temperature to 104°C to prevent condensation of the sample on the plate seal.

Step	Type	Temperature	Duration
1	Lysis	37°C	30 min.
2	Zymolase	95°C	10 min.
3	inactivation	10°C	forever
	Hold		

78 Add 100 µl of sterile ddH₂O to each well and mix.

PAUSE POINT Lysed yeast cells can be stored at –20°C for up to 1 month. Note that this can adversely affect the quality of the subsequent PCR reactions however.

Performing yeast PCR • TIMING 4 h—79 Reveal the identity of PPIs by yeast PCR by repeating Steps **12–16**, instead using HiFi Platinum Taq polymerase, DB and Term primers for pDEST-DB genes, and AD and Term primers for pDEST-AD genes in Step **12**, and adding 2 µl of template mixture from Step **78** to the distributed PCR master mix in Step **14**.

TROUBLESHOOTING

80 Send PCR products for sequencing.

PPI validation by transferring mutant entry clones into GPCA vectors by

Gateway LR reactions • TIMING 2-3 d—81 Perform LR reactions to transfer cancer mutation entry clones into GPCA vectors by repeating Steps **6–11**, instead using pSPICA-C1 or pSPICA-N2 vectors in Step **6**, and using 2 µl of cancer mutation entry clone plasmids (50–100 ng/µl) from Step **41** per well in Step **7**.

82 Verify the identity of LR reaction products by repeating Steps **12–16**, instead using pCiNeo-Univ and pCiNeo-Rev as primers in Step **12**, and empty pSPICA-C1 or pSPICA-N2 vector as a control in Step **14**. Then follow Step **17** for plasmid extraction.

TROUBLESHOOTING

Cell culture and DNA co-transfection using polyethylenimine (PEI) • TIMING

2-3 d—83 Grow human embryonic kidney (HEK) 293T cells in Dulbecco's Modified Eagle's medium (DMEM-GlutaMAX™), supplemented with 10% fetal calf serum, 100 units/ml penicillin and 100 µg/ml streptomycin at 37°C with 5% CO₂.

84 Seed 100 µl/well at a concentration of 3.0×10^4 cells per well into a desired number of 96-well plates. Incubate overnight at 37°C with 5% CO₂.

85 Prepare 100 ng of DNA per construct (pSPICA-C1 and pSPICA-N2 each fused in frame with tested protein pairs) for transfection. Add DMEM (serum and antibiotics free) to a final volume of 36 µl.

86 For every 100 samples in a 96-well plate, prepare polyethylenimine (PEI) solution by adding 225 μ l of 1 mg/ml PEI (stock solution, pH = 7.0) into 3 ml of DMEM (serum and antibiotics free). Mix well. Keep in mind to make an extra 10-15% volume to account for the media loss during pipetting, especially when using automated robotics.

87 Add 24 μ l of the PEI+DMEM mix to the wells that already have 36 μ l of DNA+DMEM. Mix well. Seal the plate with a heat sealer or aluminum foil and briefly centrifuge. Incubate for 30 min at room temperature under the tissue culture hood.

88 For DNA transfection, gently add 20 μ l of DNA+PEI+DMEM mixture to HEK293T cells. Perform two repeats. Incubate overnight at 37°C with 5% CO₂.

CRITICAL STEP Add transfection reagents over the surface of the media. Take extra care not to disrupt the cells on the bottom of the plate.

Protein interaction measurement by reading GPCA luminescence signal •

TIMING 2 h—89 Wash cells gently using 100 μ l/well of DPBS buffer containing Ca²⁺ and Mg²⁺.

90 Add 40 μ l/well of Renilla Buffer (Renilla Luciferase Assay system). Incubate for 20 min at room temperature with light shaking.

91 Prepare the Renilla luciferase substrate reagent (Renilla Luciferase Assay system) using specified buffer solution. Mix well. Add 50 μ l of substrate to each well (only handle one plate at a time).

92 Immediately measure the luciferase enzymatic activity by a Luminescence Microplate Reader (Fig. 5). The Renilla luminescence counting program has an integration time of 10 s after injection of 50 μ l of luciferase substrate reagent.

GPCA Interaction scoring and data analysis • TIMING 1 d—93 For each tested protein pair X-Y, data normalization is performed by dividing the luminescence unit of cells co-transfected with pSPICA-C1-X and pSPICA-N2-Y by the luminescence control unit either from cells co-transfected with pSPICA-C1-X and empty pSPICA-N2 vector, or from cells co-transfected with pSPICA-N2-Y and empty pSPICA-C1 vector.

94 The interaction score is defined as the mean of the log₂ of the two normalized luminescence values.

95 Titrate the “threshold” signal of quantitative outputs to an acceptable range. Set the threshold such that any pair scoring above that threshold is considered “positive” and otherwise considered “negative”. The recovery rate of false positive pairs is restricted to be 1%–2.5%.

TROUBLESHOOTING

Troubleshooting advice can be found in Table 2.

TIMING

Steps 1-2, large-scale computational design of mutation cloning primers: 1 d

Steps 3-17, preparation of mutagenesis PCR templates: 3 d

Steps 18-24, high-throughput generation of DNA fragments flanking the mutation site: 4 h

Steps 25-30, high-throughput mutagenesis of cancer variants by fusion PCR: 4 h

Steps 31-34, large-scale construction of cancer mutation entry clones: 2–3 d

Steps 35-41, verification of mutation entry clones by barcoded next-generation sequencing: 5–6 d hands-on, 2–3 weeks expansion

Steps 42-43, Gateway transfer of cancer mutations into eY2H expression vectors: 2–3 d

Steps 44-59, transforming yeast cells with DB-X or AD-Y mutant clones: 4 d

Steps 60-74, characterizing protein-protein interaction profiles of mutant clones: 5–6 d

Steps 75-78, lysis of yeast cells for PCR and sequence verification of positive interaction pairs: 2 h

Steps 79-80, performing yeast PCR: 4 h

Steps 81-82, PPI validation by transferring cancer mutant entry clones into GPCA vectors by Gateway LR reactions: 2–3 d

Steps 83-88, cell culture and DNA co-transfection using polyethylenimine (PEI): 2–3 d

Steps 89-92, protein interaction measurement by reading luminescence signal: 2 h

Steps 93-95, GPCA interaction scoring and data analysis: 1 d

ANTICIPATED RESULTS

After Gateway LR or BP reactions, the expected miniprep plasmid yield is 50–300 ng/μl in 50 μl, if the conditions are optimized (Fig. 1 and 2). The expected size range for the gene ORFs in an ideal Gateway reaction is 200–3000 bp. For barcoded next-generation sequencing, ~100 mutation pools can be generated by using 96 unique barcode adaptors (Fig. 3). Selection of sequencing platforms depends on the scale of mutagenesis. 300-400 million reads can be obtained per library using HiSeq 2500 or NextSeq 500. After low-quality bases have been trimmed, roughly 80-90% of reads can be aligned to the genes of interest. All assay systems have sensitivity limits. In general, interaction assays such as eY2H and GPCA have a sensitivity of ~20-30%. It is critical to compare the interaction capability of mutant and wild-type clones under identical experimental conditions (Fig. 4). If the mutation is located on the surface of the protein structure or at the interaction interface, it would likely impair that protein interaction. For validating interactions found by

independent assays, the fraction of positive interactions is expected to be similar to that of gold standard interactions from literature (Fig. 5).

ACKNOWLEDGEMENTS

We acknowledge the following research funds: the Cancer Prevention & Research Institute of Texas (CPRIT) grant RR160021 (N.S.); the University of Texas Systems Rising STARS award (N.S.); the NIH/NCI award number P30CA016672 (N.S.); and the University Center Foundation via the Institutional Research Grant program (to N.S.) at the University of Texas MD Anderson Cancer Center, as well as grants from National Natural Science Foundation of China (81630086, 91529305, 81427805).

REFERENCES

1. Sjoblom T et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274 (2006). [PubMed: 16959974]
2. Jones S et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321, 1801–1806 (2008). [PubMed: 18772397]
3. Parsons DW et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807–1812 (2008). [PubMed: 18772396]
4. Wood LD et al. The genomic landscapes of human breast and colorectal cancers. *Science* 318, 1108–1113 (2007). [PubMed: 17932254]
5. Fujimoto A et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* 44, 760–764 (2012). [PubMed: 22634756]
6. Gartner JJ et al. Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc. Natl. Acad. Sci. U. S. A.* 110, 13481–13486 (2013). [PubMed: 23901115]
7. Berger MF et al. The genomic complexity of primary human prostate cancer. *Nature* 470, 214–220 (2011). [PubMed: 21307934]
8. Sahni N et al. Edgotype: a fundamental link between genotype and phenotype. *Curr. Opin. Genet. Dev.* 23, 649–657 (2013). [PubMed: 24287335]
9. Garraway LA & Lander ES Lessons from the cancer genome. *Cell* 153, 17–37 (2013). [PubMed: 23540688]
10. Vidal M, Cusick ME & Barabasi AL Interactome networks and human disease. *Cell* 144, 986–998 (2011). [PubMed: 21414488]
11. Firnberg E & Ostermeier M PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One* 7, e52031 (2012). [PubMed: 23284860]
12. Fowler DM & Fields S Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807 (2014). [PubMed: 25075907]
13. Fowler DM, Stephany JJ & Fields S Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* 9, 2267–2284 (2014). [PubMed: 25167058]
14. Wei X et al. A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet.* 10, e1004819 (2014). [PubMed: 25502805]
15. Sahni N et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* 161, 647–660 (2015). [PubMed: 25910212]
16. Burrell RA, McGranahan N, Bartek J & Swanton C The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501, 338–345 (2013). [PubMed: 24048066]
17. Wang Z, Jensen MA & Zenklusen JC A Practical Guide to The Cancer Genome Atlas (TCGA). *Methods Mol. Biol.* 1418, 111–141 (2016). [PubMed: 27008012]
18. Forbes SA et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 39, D945–950 (2011). [PubMed: 20952405]
19. International Cancer Genome C et al. International network of cancer genome projects. *Nature* 464, 993–998 (2010). [PubMed: 20393554]

20. Weinberg RA Coming full circle—from endless complexity to simplicity and back again. *Cell* 157, 267–271 (2014). [PubMed: 24679541]
21. MacArthur DG et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 508, 469–476 (2014). [PubMed: 24759409]
22. Stratton MR, Campbell PJ & Futreal PA The cancer genome. *Nature* 458, 719–724 (2009). [PubMed: 19360079]
23. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615 (2011). [PubMed: 21720365]
24. del Sol A, Balling R, Hood L & Galas D Diseases as network perturbations. *Curr. Opin. Biotechnol.* 21, 566–571 (2010). [PubMed: 20709523]
25. Ryan CJ et al. High-resolution network biology: connecting sequence with function. *Nat. Rev. Genet.* 14, 865–879 (2013). [PubMed: 24197012]
26. Li Y, Sahni N & Yi S Comparative analysis of protein interactome networks prioritizes candidate genes with cancer signatures. *Oncotarget* 7:78841–78849. (2016). [PubMed: 27791983]
27. Kim E et al. Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov* 6, 714–726 (2016). [PubMed: 27147599]
28. Cheung LW et al. Naturally occurring neomorphic PIK3R1 mutations activate the MAPK pathway, dictating therapeutic response to MAPK pathway inhibitors. *Cancer Cell* 26, 479–494 (2014). [PubMed: 25284480]
29. Mali P et al. RNA-guided human genome engineering via Cas9. *Science* 339, 823–826 (2013). [PubMed: 23287722]
30. Cong L et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823 (2013). [PubMed: 23287718]
31. Jinek M et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821 (2012). [PubMed: 22745249]
32. Fire A et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811 (1998). [PubMed: 9486653]
33. Petersen CP, Bordeleau ME, Pelletier J & Sharp PA Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* 21, 533–542 (2006). [PubMed: 16483934]
34. Platt RJ et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455 (2014). [PubMed: 25263330]
35. Wang X et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 30, 159–164 (2012). [PubMed: 22252508]
36. Zhong Q et al. Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* 5, 321 (2009). [PubMed: 19888216]
37. Chakravarti A, Clark AG & Mootha VK Distilling pathophysiology from complex disease genetics. *Cell* 155, 21–26 (2013). [PubMed: 24074858]
38. Lin S, Yin YA, Jiang X, Sahni N & Yi S Multi-OMICs and Genome Editing Perspectives on Liver Cancer Signaling Networks. *BioMed research international* 2016, 6186281 (2016). [PubMed: 27403431]
39. Fuxman Bass JI et al. Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* 161, 661–673 (2015). [PubMed: 25910213]
40. Barrera LA et al. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science* 351, 1450–1454 (2016). [PubMed: 27013732]
41. Taipale M et al. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* 150, 987–1001 (2012). [PubMed: 22939624]
42. Hartley JL, Temple GF & Brasch MA DNA cloning using in vitro site-specific recombination. *Genome Res.* 10, 1788–1795 (2000). [PubMed: 11076863]
43. Yang X et al. A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* 8, 659–661 (2011). [PubMed: 21706014]
44. Yachie N et al. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* 12, 863 (2016). [PubMed: 27107012]

45. Rolland T et al. A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226 (2014). [PubMed: 25416956]
46. Rual JF et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173–1178 (2005). [PubMed: 16189514]
47. Dreze M et al. ‘Edgetic’ perturbation of a *Caenorhabditis elegans* BCL-2 ortholog. *Nat. Methods* 6, 843–849 (2009). [PubMed: 19855391]
48. Rual JF et al. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* 14, 2128–2135 (2004). [PubMed: 15489335]
49. Boxem M et al. A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* 134, 534–545 (2008). [PubMed: 18692475]
50. Fields S & Song O A novel genetic system to detect protein-protein interactions. *Nature* 340, 245–246 (1989). [PubMed: 2547163]
51. Dreze M et al. High-quality binary interactome mapping. *Methods Enzymol.* 470, 281–315 (2010). [PubMed: 20946815]
52. Yu H et al. Next-generation sequencing to generate interactome datasets. *Nat. Methods* 8, 478–480 (2011). [PubMed: 21516116]
53. Walhout AJ & Vidal M A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* 9, 1128–1134 (1999). [PubMed: 10568752]
54. Cassonnet P et al. Benchmarking a luciferase complementation assay for detecting protein complexes. *Nat. Methods* 8, 990–992 (2011). [PubMed: 22127214]
55. Remy I & Michnick SW A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat. Methods* 3, 977–979 (2006). [PubMed: 17099704]
56. O’Halloran DM PrimerMapper: high throughput primer design and graphical assembly for PCR and SNP detection. *Scientific reports* 6, 20631 (2016). [PubMed: 26853558]

Figure 1. Functional variomics pipeline to characterize cancer mutations

A long catalogue of human cancer mutations (including missense mutations) is available from next-generation sequencing projects (TCGA, 1000 genomes, dbGAP, etc). We provide a functional variomics platform to gain insights into the function of these mutations. In order to do this, one needs to first generate and express these mutant clones in a high-throughput manner, as detailed in Steps 1-41. Large-scale functional variomics platforms have recently emerged to characterize cancer mutations, including but not limited to high-throughput enhanced yeast two-hybrid (HT-eY2H) assay (Steps 42-80) and mammalian cell-based *Gaussia princeps* protein-fragment complementation assay (GPCA; Steps 81-95). HT-eY2H enables characterization of protein interactions for large numbers of samples based on the reconstitution of the yeast Gal4 transcription factor from activation domain (AD) and DNA-binding domain (DB). GPCA detects protein-protein interactions by complementation of

Gaussia princeps luciferase (represented in yellow) from the two fragments of the luciferase hGLuc F[1] and hGLuc F[2] (labeled by arrows). WT, wild-type. Stars represent mutations; Node represents protein; Edge represents interaction between proteins.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. High-throughput construction of a cancer mutation library

Starting from ORFs that are available within the ORFeome collection, two separate primary PCRs are performed. For the primary PCRs, two universal primers, AD-Tag1 and Term-Tag2 (sequences shown in Table 1), and two mutation-specific internal forward and reverse primers are employed. The two universal primers allow the preservation of the attB sites (shown by green boxes) on both ends of the gene. The mutation-specific primers, MutF and MutR, encompassing the desired mutation, are designed to have an overlapping region of ~40 base pairs. The two DNA fragments flanking the mutation of a cancer gene are amplified using the primer pair AD-Tag1 and MutR, and the primer pair Term-Tag2 and MutF, respectively. For the fusion PCR, the two primary PCR products are fused together using the primer pair Tag1 and Tag2 (sequences shown in Table 1) to generate the mutant allele. DNA fragments during the mutagenesis process are analyzed on 1% Agarose E-Gel (96 samples). Fragment I, Fragment II, and fusion PCR products are shown in a 96-well format (left, middle and right, respectively). As is shown, fusion PCR products exhibit DNA sizes as an addition of the two fragment sizes.

Figure 3. Confirmation of mutation entry clones by barcoded next-gen sequencing

Products of the PCR mutagenesis reaction are cloned into the pDONR223 vector via Gateway BP reaction and transformed into competent DH5 α bacterial cells. Transformants are spotted on LB + Spectinomycin plates to isolate single colonies. One plate is shown here as a representation. This scheme illustrates the steps involved in mutation sample preparation and next-generation sequencing confirmation. PCR amplification should be done for each mutant at the single colony level prior to sample preparation (illustrated by 96 well plates M1-M4). If a given gene has multiple mutations to clone (*e.g.*, four as shown), these mutation PCR products should be separated into different pools (Steps 35-36). A unique barcode adaptor is added to each pool by ligation (Step 37). All pooled mutation PCR products are then sequence-verified by next-generation sequencing technologies (454-FLX, Illumina Solexa, etc) (Steps 38-41).

Figure 4. Characterizing cancer mutations by high-throughput enhanced yeast two-hybrid (HT-eY2H) assay

Mutant and wild-type genes are transformed and expressed in yeast cells, and plated on selective plates. In a large-scale experiment, mutants (M1, M2 and M1') and corresponding wild-type counterparts (X and X') are compared on the same media plate condition (top panel), in terms of their ability to form protein interactions. For the HT-eY2H assay, DB-X and AD-Y fusion proteins are transformed into yeast Y8930 and Y8800 respectively, and grown in SC-Leu (SC-L) and SC-Trp (SC-W) media respectively. These yeast strains are then mated in YPD media, and undergo diploid selection in SC-Leu-Trp (SC-LW) media for

successful mating events. Finally, protein interactions are identified on SC-LWH+3AT (1 mM 3AT) media based on the reporter *HIS3* expression. A stringent control must be performed in SC-LH+3AT+CHX (cycloheximide, 1 mg/l) to detect ‘autoactivators’ for cells that have lost the AD-Y plasmid. Protein-protein interactions are tested in a pairwise fashion, and four times independently for activation of *HIS3* reporter in a *CHX* sensitive manner. Representative samples are shown. Protein pairs that score positive three or four times are considered genuine Y2H interactors. Control 1 expresses DB and AD plasmids without any fusion. Control 2 expresses DB-pRB and AD-E2F1 fusion proteins, forming an interaction, and is CHX sensitive. Controls 3, 4, and 5 express DB-Fos and AD-Jun, DB-GAL4 and AD, and DB-DP and AD-E2F1, respectively, all of which exhibit positive interactions but are CHX resistant. Control 6 expresses DB-DP and AD-E2F1, and is CHX sensitive. This test result shows that Gene A mutant is defective, while Gene B mutant is not for their respective interaction test compared to wild-type counterparts. X, Y, protein interaction partners. M, mutation. DB, DNA-binding domain of Gal4. AD, activation domain of Gal4. SC, synthetic complete media for yeast.

Figure 5. Characterizing cancer mutations by *Gaussia princeps* luciferase protein-fragment complementation assay (GPCA)

To profile interaction changes by GPCA, mutant and wild-type Entry clones of Gene A are transferred into the two GPCA vectors (shown in green and pink, respectively), expressing hGLuc F[1]-A and hGLuc F[2]-interactor fusion proteins in human cells. Black bar and red bar represent Gene A and its interactor gene, respectively. Star represents mutation. Pairs of vectors expressing hGLuc F[1]-A and hGLuc F[2]-interactor are co-transfected into cells at the same time. Interaction profiling of mutants and wild-type counterparts is performed in replicates. Signal intensities are read by a 96-well plate reader, and then processed for data analysis. Overall the repeats are consistent across replicates. A panel of GPCA results with two repeats is shown on the right side of the figure.

TABLE 1 |**Primers for mutagenesis and sequence confirmation****Primers for high-throughput mutagenesis**

AD-Tag1: 5'-GGCAGACGTGCCTCACTACTCGCGTTTGAATCACTACAGGG-3' (Step 19)

Term-Tag2: 5'-CTGAGCTTGACGCATTGCTAGGAGACTTGACCAAACCTCTGGCG-3' (Step 20)

Tag1: 5'-GGCAGACGTGCCTCACTACT-3' (Step 26)

Tag2: 5'-CTGAGCTTGACGCATTGCTA-3' (Step 26)

M13F: 5'-CCCAGTCACGACGTTGTAAAACG-3' (Step 34)

M13R: 5'-GTGTCTCAAAATCTCTGATGTTAC-3' (Step 34)

Primers for mutation confirmation in HT-eY2H

AD: 5'-CGCGTTTGAATCACTACAGGG-3' (Step 12, 43, 79)

DB: 5'-GGCTTCAGTGGAGACTGATATGCCTC-3' (Step 43, 79)

Term: 5'-GGAGACTTGACCAAACCTCTGGCG-3' (Step 12, 43, 79)

Primers for mutation confirmation in GPCA

pCiNeo-Univ: 5'-CAGCTCTTAAGGCTAGAGTAC-3' (Step 82)

pCiNeo-Rev: 5'-CACTGCATTCTAGTTGTGGTTGTCC-3' (Step 82)

Table 2 |

Troubleshooting table

Step	Problem	Possible reason	Solution
15, 23, 29, 34, 43, 79 and 82	Significant loss of PCR reaction volume	Insufficient sealing	Aluminum foil is the easiest method to sealing plates, but may lead to evaporation. Alternatively, a robotic heat sealer or strip-caps may be used, both of which offer superior sealing capability.
30	Insufficient quality of the PCR amplification	Experimental conditions not optimized	Based on empirical testing, we obtained optimal PCR results using a PCR cycle number of 30, which is lower than the usual 33-35 cycles. In addition, the amount of Mg^{2+} can be reduced to increase PCR specificity. Furthermore, using a temperature of 68°C for the KOD hot start elongation step could also often enhance the quality.
56	Too few yeast transformants	Yeast cells not fresh or grown to an incorrect density	First make sure that the yeast cells used to inoculate the starter culture of YPD were grown fresh on a plate, and that the final optical density at 600 nm was between 0.6 and 1.0. Try transforming a well-purified (maxi prep) plasmid to check if the problem lies in the material that is transformed. If this is the case, increase the amount of plasmid and/or PCR product that is transformed.
71	Neighboring colonies merge on the agar plate	Liquid has a tendency to spread out	Reduce the volume spotted on agar, or dry the agar plates for 2 weeks at room temperature.
79	Smeary PCR products	Yeast lysis not thorough	Using fresh yeast cells from overnight growth is the most important component of a successful PCR reaction. The most consistent results are obtained when the lysed yeast cells are added to the PCR mixture without delay. Finally, a longer extension time (<i>e.g.</i> , 4 min) than would normally be appropriate for inserts of a given size results in improved PCR results.