ONLINE SUPPLEMENTARY MATERIAL FOR

Edgetic perturbation models of human inherited disorders

Quan Zhong[1,*] Nicolas Simonis[1,*], Qian-Ru Li[1,*], Benoit Charloteaux[1,2,*], Fabien Heuze[1,2,*], Niels Klitgord[1,*], Stanley Tam[1], Haiyuan Yu[1], Kavitha Venkatesan[1], Danny Mou[1], Venus Swearingen, Muhammed A Yildirim[1], Han Yan[1], Amélie Dricot[1], David Szeto[1], Chenwei Lin[1], Tong Hao[1], Changyu Fan[1], Stuart Milstein[1], Denis Dupuy[1], Robert Brasseur[2], David E Hill[1], Michael E Cusick[1] & Marc Vidal[1]

[1]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]Centre de Biophysique Moléculaire Numérique, Faculté Universitaire des Sciences Agronomiques de Gembloux, 2 Passage des Déportés, B-5030 Gembloux, Wallonia, Belgium. *These authors contributed equally to this work. Correspondence should be addressed to M.V. (marc_vidal@dfci.harvard.edu).

This PDF includes:
Supplementary Text (Text, Reference, Supplementary Figure legends)
Supplementary Figures (Figure S1-9)
Supplementary Tables (Table S1-4)

SUPPLEMENTARY TEXT

CBS (cystathionine β-synthase)

Enzyme deficiency caused by mutations in *CBS* gives rise to the metabolic disorder Homocystinuria. Five disease-causing mutant alleles from HGMD, distributed along the coding sequence of CBS, were cloned and tested for interactions against three interactors of the respective wild-type protein (Rual *et al.*, 2005) (Figure S2A). Several classes of interaction-defective alleles were identified. One allele (I278T) behaved as a null, eliminating all three interactions. Two (P49L and P422L) behaved as "pseudo-wild-type", retaining all three protein-protein interactions. The other two alleles (P145L and L539S) have one Y2H interaction retained while two other interactions lost.

Three-dimensional structural information is available for CBS (Meier *et al.*, 2001). The null-like allele (I278T) has non-conservative amino acid substitutions in the highly structured internal region of the protein, possibly grossly disrupting protein conformation. The allele with partial loss of protein interaction (P145L) and the wild-type-like allele (P49L) are on the surface (Figure S2B). Pro49 locates near the heme-binding residues Cys52 and His65. Given the unique conformational rigidity of proline, the P49L mutation might affect enzyme function via perturbation of the heme-binding pocket.

The P422L mutant allele fully preserves enzyme activity (Figure S2D), consistent with its wild-type-like Y2H interaction profile. Both the null-like I278T allele and the edgetic P145L allele exhibit nearly complete loss of enzyme activity (Figure S2D). The former (I278T) likely results from grossly disrupt

protein, while the latter (P145L) is likely caused by defective dimerization as observed in the Y2H analysis.

Allele-specific perturbations of CBS mutant proteins were found to be associated with a treatment response. Pyridoxine, a precursor of the CBS cofactor pyridoxal phosphate (PLP), alleviates CBS deficiency and reduces the associated disease symptoms (Mudd *et al.*, 1985). Not all patients respond to pyridoxine. A patient with the wild-type-like P422L allele is not pyridoxine responsive (Maclean *et al.*, 2002), suggesting that increased PLP cofactor does not enhance CBS enzyme activity of alleles with presumably largely unaffected structure. Patients carry the alleles with partial loss of protein interactions, P145L (Kozich *et al.*, 1993) or L539S (Aral *et al.*, 1997), are pyridoxine responsive, consistent with partially perturbed protein structures that may be stabilized by cofactor binding. Some patients with the null-like I278T allele are reportedly pyridoxine responsive (Kluijtmans *et al.*, 1999). However, it was recently suggested that the mechanism of pyridoxine responsiveness in I278T patients may involve mechanisms besides direct enhancement of CBS activity by PLP (Chen *et al.*, 2006). The null-like interaction profile of I278T also supports a grossly altered structure, unlikely to be directly restored by cofactor binding.

HGD (homogenistate dioxygenase)

Mutations in *HGD* give rise to Alkaptonuria, a benign inborn error of metabolism characterized by excess homogentisic acid in body fluids, causing several symptoms including "black urine" (Garrod, 1902; Phornphutkul *et al.*, 2002). Five

disease-causing mutant alleles from HGMD distributed along the coding sequence of HGD were cloned and tested for interactions against three interactors of the respective wild-type protein identified in our screen (HGD, NUDT18, NIF3L1) (Figure S3A). We found a similar distribution of allele classes: alleles that are null with respect to all tested protein-protein interactions (V300G and E42A), a pseudo-wild-type allele (H371R), and alleles causing interaction-specific perturbation to HGD (R225H and L25P) (Figure S3A).

Three-dimensional structure is available for HGD (Titus *et al.*, 2000). Similar to CBS, the null-like allele (V300G) also has non-conservative amino acid substitutions in the highly structured internal regions of the protein (Figure S3B), possibly grossly disrupting protein conformation. The other null allele (E42A) likely causes defects in HGD oligomerization, since the E42 residue forms a salt-bridge with the three-fold related Arg336 in the HGD hexamer. Both null alleles (V300G and E42A) have dramatically loss of enzyme activity (Figure S3D). Alleles causing interaction-specific defects (R225H and L25P) correspond to mutated residues on the surface of the protein (Figure S3B). The Arg225 and Leu25 residues participate in the dimerization of HGD trimers. Substitution of these residues by His and Pro, respectively, should not affect trimerization of the hexamer, given the nature of the substitutions and of the two-fold interface of HGD oligomer. The edgetic allele (R225H) with nearly complete loss of enzyme activity (Figure S3D) showed clearly reduced dimerization in the Y2H analysis (Figure S3A), possibly leading to defective hexamer formation. The pseudo-wild-type allele (H371R) also has mutated residues on the surface (Figure S3B). The

side chain of His371 directly coordinates the cofactor Fe^{2+}. Therefore the H371R mutant allele is likely defective for ligand binding leading to loss of enzyme activity (Figure S3D) (Rodriguez *et al.*, 2000).

ACTG1 *(*cytoplasmic γ-actin)

Five mutations in *ACTG1* are associated with progressive, sensorineural hearing loss as annotated in HGMD. Profiling of deafness-associated ACTG1 alleles revealed wild-type-like and interaction-specific perturbations. Three ACTG1 alleles (T89I, K118M, and T278I) retain all wild-type interactions (Figure S4A). Of these, Thr89 and Thr278 are completely buried in the structure of the ortholog bovine β-actin (Chik *et al.*, 1996), and the K118M substitution modifies the physicochemical properties of the solvent-exposed Lys118 residue (Figure S4B). The lack of interaction defects in these disease-causing alleles may simply result from the few interactions that were analyzed. Thr89 is implicated in an interaction with an actin bundling protein, fimbrin (Adams and Botstein, 1989), but this interaction was not tested here. Two mutations (P264L and P332A) show severe interaction defects with actin depolymerizing factors (CFL1, CFL2, and DSTN) (Figure S4A), suggesting defective actin dynamics leading to hearing loss. Pro264 is buried in the bovine β-actin structure (Figure S4B), mutation of which likely results in structural alterations. Pro332 is surface exposed (Figure S4B) and falls into one of the five biochemically identified peptide stretches (328-338) that mediate binding of actin to cofilin (Mannherz *et al.*, 2007). All five ACTG1 mutant alleles preserve interactions with wild-type β-actin (ACTB) and γ-actin

(ACTG1) (Figure S4A), consistent with possible incorporation of mutant monomers in actin filaments *in vivo*, causing dominant negative effects.

CDK4 (cyclin dependent kinase 4)

CDK4 controls cell cycle progression and is negatively regulated by cyclin-dependent kinase inhibitors (Ortega *et al.*, 2002). Four *CDK4* germline mutations are annotated in HGMD with increased risk of melanoma. All four CDK4 mutant alleles were cloned and tested for interactions against three previously identified interactors (Rual *et al.*, 2005), and another known CDK4 inhibitor available in our human ORFeome collection (Lamesch *et al.*, 2007), CDKN2C (cyclin-dependent kinase inhibitor 2C). Both R24H and R24C mutant alleles show reduced interaction with CDKN2C (Figure S5A), consistent with available crystallographic data and oncogenic activation of CDK4. The other two mutations (N41S and S52N) affect residues in the vicinity of cyclin binding site but do not make direct contact with D-type cyclin in the crystal structures (Day *et al.*, 2009) (Figure S5B). N41S was found in the germline of a patient with no family history of melanoma (Guldberg *et al.*, 1997). S52N was found in a family with a history of melanoma but this mutation is not carried by all affected individuals (Holland *et al.*, 1999). The conservative nature of these mutations and their unaffected interaction profile in Y2H strongly suggest that the pathological relevance for these two alleles is uncertain. To test the possibility that CDK4 mutant alleles may gain new interactions, we carried out Y2H screens for both the wild-type and four mutant CDK4 proteins against a set of ~12,200 human open reading frames

(Lamesch *et al.*, 2007). However, at this stage we did not recover any gain-of-interaction for any of the four mutant CDK4 proteins yet.

PRKAR1A (cAMP-dependent protein kinase type Iα regulatory subunit)

Mutations in *PRKAR1A* are associated with Carney complex, a multiple neoplasia syndrome. About 90% of PRKAR1A mutations recorded in HGMD result in premature stop codons or frameshifts (9 nonsense, 1 missense, 4 in-frame and 26 out-of-frame insertions or deletions, and 12 splice site mutations). PRKAR1A regulates one of the two types of PKA, type I PKA. Reduced PRKAR1A levels likely perturb the balance between type I and type II PKA, leading to abnormal cell growth and proliferation (Stergiopoulos and Stratakis, 2003).

We cloned the one missense and nine nonsense PRKAR1A mutant alleles annotated in HGMD and tested them for interactions against three interactors of the wild-type PRKAR1A (Rual *et al.*, 2005). All fusion proteins carrying nonsense mutations exhibited dramatically reduced expression in yeast, whereas the missense R74C allele is normally expressed (Figure S6A), resembling the expression pattern of mutant PRKAR1A observed in patients (Kirschner *et al.*, 2000). Despite reduced expression, most mutant proteins preserved one or two wild-type interactions, except for Q28X and Q37X (Figure S6A). The two null-like alleles (Q28X and Q37X) truncate the N-terminal RIIα domain of PRKAR1A (Figure S6B). The R42X mutant allele, located between the two helices of the helix-turn-helix RIIα domain (Banky *et al.*, 2003), loses interactions with AKAP10

and PLEKHF2 but preserves the interaction with MGC13057 (Figure S6A). The expressed missense allele R74C and three nonsense alleles (K63X, Q304X and S307X) preserve the interaction with AKAP10 (Figure S6A), a kinase anchoring protein that binds both type I and type II PKA (Huang *et al.*, 1997b). This dual specificity of AKAP10 allows differential targeting of type I and type II PKA, crucial for integrated signaling (Huang *et al.*, 1997a). The interaction of PRKAR1A mutant proteins with AKAP10 could interfere with the binding of functional PKA to AKAP10 and cause additional imbalance between type I and type II PKA. Such an interaction could account for more severe phenotypes reported for patients with expressed mutant PRKAR1A (Horvath *et al.*, 2008).

REFERENCES

Adams, A. E., and Botstein, D. (1989). Dominant suppressors of yeast actin mutations that are reciprocally suppressed. Genetics *121*, 675-683.

Aral, B., Coude, M., London, J., Aupetit, J., Chasse, J. F., Zabot, M. T., Chadefaux-Vekemans, B., and Kamoun, P. (1997). Two novel mutations (K384E and L539S) in the C-terminal moiety of the cystathionine β-synthase protein in two French pyridoxine-responsive homocystinuria patients. Hum Mutat *9*, 81-82.

Banky, P., Roy, M., Newlon, M. G., Morikis, D., Haste, N. M., Taylor, S. S., and Jennings, P. A. (2003). Related protein-protein interaction modules present drastically different surface topographies despite a conserved helical platform. J Mol Biol *330*, 1117-1129.

Chen, X., Wang, L., Fazlieva, R., and Kruger, W. D. (2006). Contrasting behaviors of mutant cystathionine beta-synthase enzymes associated with pyridoxine response. Hum Mutat *27*, 474-482.

Chik, J. K., Lindberg, U., and Schutt, C. E. (1996). The structure of an open state of β-actin at 2.65 Å resolution. J Mol Biol *263*, 607-623.

Day, P. J., Cleasby, A., Tickle, I. J., O'Reilly, M., Coyle, J. E., Holding, F. P., McMenamin, R. L., Yon, J., Chopra, R., Lengauer, C., and Jhoti, H. (2009). Crystal structure of human CDK4 in complex with a D-type cyclin. Proc Natl Acad Sci USA *106*, 4166-4170.

Garrod, A. E. (1902). The incidence of alkaptonuria: a study in chemical individuality. *Lancet ii*, 1616-1620.

Guldberg, P., Kirkin, A. F., Gronbaek, K., thor Straten, P., Ahrenkiel, V., and Zeuthen, J. (1997). Complete scanning of the *CDK4* gene by denaturing gradient gel electrophoresis: a novel missense mutation but low overall frequency of mutations in sporadic metastatic malignant melanoma. Int J Cancer *72*, 780-783.

Holland, E. A., Schmid, H., Kefford, R. F., and Mann, G. J. (1999). *CDKN2A* (P16[INK4a]) and *CDK4* mutation analysis in 131 Australian melanoma probands: effect of family history and multiple primary melanomas. Genes Chromosomes Cancer *25*, 339-348.

Horvath, A., Bossis, I., Giatzakis, C., Levine, E., Weinberg, F., Meoli, E., Robinson-White, A., Siegel, J., Soni, P., Groussin, L., Matyakhina, L., Verma, S., Remmers, E., Nesterova, M., Carney, J. A., Bertherat, J., and Stratakis, C. A. (2008). Large deletions of the *PRKAR1A* gene in Carney complex. Clin Cancer Res *14*, 388-395.

Huang, L. J., Durick, K., Weiner, J. A., Chun, J., and Taylor, S. S. (1997a). D-AKAP2, a novel protein kinase A anchoring protein with a putative RGS domain. Proc Natl Acad Sci USA *94*, 11184-11189.

Huang, L. J., Durick, K., Weiner, J. A., Chun, J., and Taylor, S. S. (1997b). Identification of a novel protein kinase A anchoring protein that binds both type I and type II regulatory subunits. J Biol Chem *272*, 8057-8064.

Jeffrey, P. D., Tong, L., and Pavletich, N. P. (2000). Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. Genes Dev *14*, 3115-3125.

Kirschner, L. S., Sandrini, F., Monbo, J., Lin, J. P., Carney, J. A., and Stratakis, C. A. (2000). Genetic heterogeneity and spectrum of mutations of the *PRKAR1A* gene in patients with the carney complex. Hum Mol Genet *9*, 3037-3046.

Kluijtmans, L. A., Boers, G. H., Kraus, J. P., van den Heuvel, L. P., Cruysberg, J. R., Trijbels, F. J., and Blom, H. J. (1999). The molecular basis of cystathionine β-synthase deficiency in Dutch patients with homocystinuria: effect of CBS genotype on biochemical and clinical phenotype and on response to treatment. Am J Hum Genet *65*, 59-67.

Kozich, V., de Franchis, R., and Kraus, J. P. (1993). Molecular defect in a patient with pyridoxine-responsive homocystinuria. Hum Mol Genet *2*, 815-816.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., Rogers, J., Lawlor, S., McLaren, S., Dricot, A., Borick, H., Cusick, M. E., Vandenhaute, J., Dunham, I., Hill, D. E., and Vidal, M. (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307-315.

Maclean, K. N., Gaustadnes, M., Oliveriusova, J., Janosik, M., Kraus, E., Kozich, V., Kery, V., Skovby, F., Rudiger, N., Ingerslev, J., Stabler, S. P., Allen, R. H., and Kraus, J. P. (2002). High homocysteine and thrombosis without connective tissue disorders are associated with a novel class of cystathionine β-synthase (CBS) mutations. Hum Mutat *19*, 641-655.

Mannherz, H. G., Ballweber, E., Galla, M., Villard, S., Granier, C., Steegborn, C., Schmidtmann, A., Jaquet, K., Pope, B., and Weeds, A. G. (2007). Mapping the ADF/cofilin binding site on monomeric actin by competitive cross-linking and peptide array: evidence for a second binding site on monomeric actin. J Mol Biol *366*, 745-755.

Meier, M., Janosik, M., Kery, V., Kraus, J. P., and Burkhard, P. (2001). Structure of human cystathionine β-synthase: a unique pyridoxal 5'-phosphate-dependent heme protein. EMBO J *20*, 3910-3916.

Mudd, S. H., Skovby, F., Levy, H. L., Pettigrew, K. D., Wilcken, B., Pyeritz, R. E., Andria, G., Boers, G. H., Bromberg, I. L., Cerone, R., and et al. (1985). The natural history of homocystinuria due to cystathionine β-synthase deficiency. Am J Hum Genet *37*, 1-31.

Ortega, S., Malumbres, M., and Barbacid, M. (2002). Cyclin D-dependent kinases, INK4 inhibitors and cancer. Biochim Biophys Acta *1602*, 73-87.

Phornphutkul, C., Introne, W. J., Perry, M. B., Bernardini, I., Murphey, M. D., Fitzpatrick, D. L., Anderson, P. D., Huizing, M., Anikster, Y., Gerber, L. H., and Gahl, W. A. (2002). Natural history of alkaptonuria. N Engl J Med *347*, 2111-2121.

Rodriguez, J. M., Timm, D. E., Titus, G. P., Beltran-Valero De Bernabe, D., Criado, O., Mueller, H. A., Rodriguez De Cordoba, S., and Penalva, M. A. (2000). Structural and functional analysis of mutations in alkaptonuria. Hum Mol Genet *9*, 2341-2350.

Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L.

V., Wong, S. L., Franklin, G., Li, S.*, et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Stergiopoulos, S. G., and Stratakis, C. A. (2003). Human tumors associated with Carney complex and germline *PRKAR1A* mutations: a protein kinase A disease! FEBS Lett *546*, 59-64.

Titus, G. P., Mueller, H. A., Burgner, J., Rodriguez De Cordoba, S., Penalva, M. A., and Timm, D. E. (2000). Crystal structure of human homogentisate dioxygenase. Nat Struct Biol *7*, 542-546.

Walhout, A. J., and Vidal, M. (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods *24*, 297-306.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1 Distribution of autosomal dominant and autosomal recessive diseases with respect to their associated "in-frame" mutations, excluding all human orthologs of essential genes in mouse, fly, worm and yeast. Mutations in each gene associated with each mode of inheritance are grouped as one trait. Each data point represents the percentage of either autosomal recessive (blue bar) or autosomal dominant (red bar) traits that have a fraction of "in-frame" mutations no less than the value on the X axis. Statistical significance of the observed difference between distributions is assessed by Mann-whitney U test ($P < 4.0 \times 10^{-13}$). The number of traits, genes, diseases and total mutations in each bin are provided in Supplementary table 1.

Figure S2 Profiling disease-causing allele-specific interaction defects of CBS (cystathione β-synthase). (**A**) Y2H analysis and immunoblotting detection of wild-type and mutant proteins of CBS. (**B**) Residues affected by mutations are shown on the reported CBS structure (Meier *et al.*, 2001), with relative solvent-accessible surface area (%ASA) indicated for surface exposed residues. (**C**) Vertical lines along a Pfam domain representation of CBS show positions of mutations. (**D**) Score of each Y2H interaction. Activation of at least two of the three reporter genes was taken as a positive interaction. Interaction pairs showing less than two positive reporters are scored as "-". Interaction pairs showing the same number of positive reporters as the corresponding wild type are scored as "+". Interactions that lose expression of one reporter but still show

expression of the other two reporters are scored as "R". Biochemical activities for P422L (Maclean *et al.*, 2002) and I278T (Chen *et al.*, 2006) were obtained from the literature. The activity of CBS P145L was obtained from the CBS mutation database (http://www.uchsc.edu/cbs/cbsdata/cbsmain.htm).

Figure S3 Profiling disease-causing allele-specific interaction defects of HGD (homogenistate dioxygenase). (**A**) Y2H analysis and immunoblotting detection of wild-type and mutant proteins of HGD. (**B**) Residues affected by mutations are shown on the reported HGD structure (Titus *et al.*, 2000), with relative solvent-accessible surface area (%ASA) indicated for surface exposed residues. (**C**) Vertical lines along a Pfam domain representation of HGD show positions of mutations. (**D**) Score of each Y2H interaction. Activation of at least two of the three reporter genes was taken as a positive interaction. Interaction pairs showing less than two positive reporters are scored as "-". Interaction pairs showing the same number of positive reporters as the corresponding wild type are scored as "+". Interactions that lose expression of one reporter but still show expression of the other two reporters are scored as "R". Biochemical activities for H371R, V300G, R225H, and E42A were obtained from the literature (Rodriguez *et al.*, 2000).

Figure S4 Profiling disease-causing allele-specific interaction defects of ACTG1 (cytoplasmic γ-actin). (**A**) Y2H analysis and immunoblotting detection of wild-type and mutant proteins of ACTG1. (**B**) Residues affected by mutations are shown

on the structure (Chik *et al.*, 1996), with relative solvent-accessible surface area (%ASA) indicated for surface exposed residues. (**C**) Vertical lines along a Pfam domain representation of ACTG1 show positions of mutations. (**D**) Score of each Y2H interaction. Activation of at least two of the three reporter genes was taken as a positive interaction. Interaction pairs showing less than two positive reporters are scored as "-". Interaction pairs showing the same number of positive reporters as the corresponding wild type are scored as "+". Interactions that lose expression of one reporter but still show expression of the other two reporters are scored as "R".

Figure S5 Profiling disease-causing allele-specific interaction defects of CDK4 (cyclin dependent kinase 4). (**A**) Y2H analysis and immunoblotting detection of wild-type and mutant proteins of CDK4. (**B**) Residues affected by mutations are shown on the structure of CDK4 (in grey) in complex with CCND3 (Day *et al.*, 2009) (in green) with relative solvent-accessible surface area (%ASA) shown. Location of the CDKN2C inhibitor was extrapolated from the structure (Jeffrey *et al.*, 2000) of CDK6 (not shown) in complex with CDKN2C (in light blue). Residues (D67 and D76) which make salt bridges with CDK4 (R24) are shown as sticks. (**C**) Vertical lines along a Pfam domain representation of CDK4 show positions of mutations. (**D**) Score of each Y2H interaction. Activation of at least two of the three reporter genes was taken as a positive interaction. Interaction pairs showing less than two positive reporters are scored as "-". Interaction pairs showing the same number of positive reporters as the corresponding wild type

are scored as "+". Interactions that lose expression of one reporter but still show expression of the other two reporters are scored as "R".

Figure S6 Profiling disease-causing allele-specific interaction defects of PRKAR1A (cAMP-dependent protein kinase type Iα regulatory subunit). (**A**) Y2H analysis and immunoblotting detection of wild-type and mutant proteins of PRKAR1A. Proteins with expected sizes are labeled with green rectangles. (**B**) Vertical lines along a Pfam domain representation of PRKAR1A show positions of mutations. (**C**) Score of each Y2H interaction. Activation of at least two of the three reporter genes was taken as a positive interaction. Interaction pairs showing less than two positive reporters are scored as "-". Interaction pairs showing the same number of positive reporters as the corresponding wild type are scored as "+". Interactions that lose expression of one reporter but still show expression of the other two reporters are scored as "R".

Figure S7 Distribution of residues affected by disease-causing missense mutations with respect to their relative solvent-accessible surface area in X-ray protein structures. Error bars represent standard errors of the mean.

Figure S8 Average fractions of "in-frame" mutations per gene associated with either autosomal dominant or autosomal recessive disease. The *P*-value of the observed difference, measured by Mann-Whitney U test, is shown.

Figure S9 Phenotypes and interaction score of the five controls of Y2H assay
(Walhout and Vidal, 2001) are show.

Figure S1

A

	CBS	P49L	P145L	I278T	P422L	L539S	
β-Gal assay							CBS
							C6orf55
							FXR2
SC-HIS							CBS
							C6orf55
							FXR2
SC-URA							CBS
							C6orf55
							FXR2

Anti-Gal4

Anti-a-tubulin

B

P145 (15%) PLP P49 (81%) Heme I278

C

P49L   P145L   I278T   P422L   L539S

| | PALP | | CBS | | 551 |
80   376   416   512
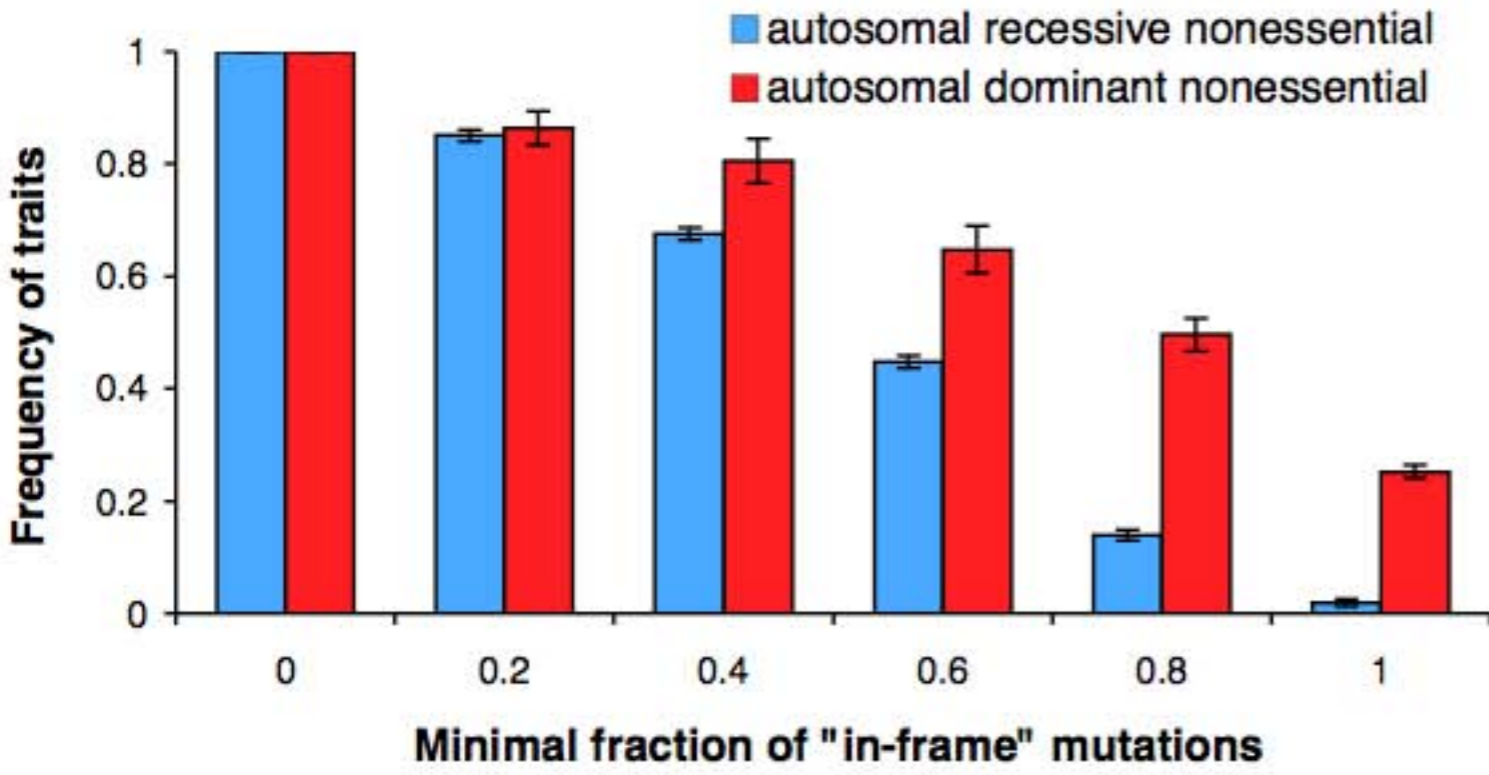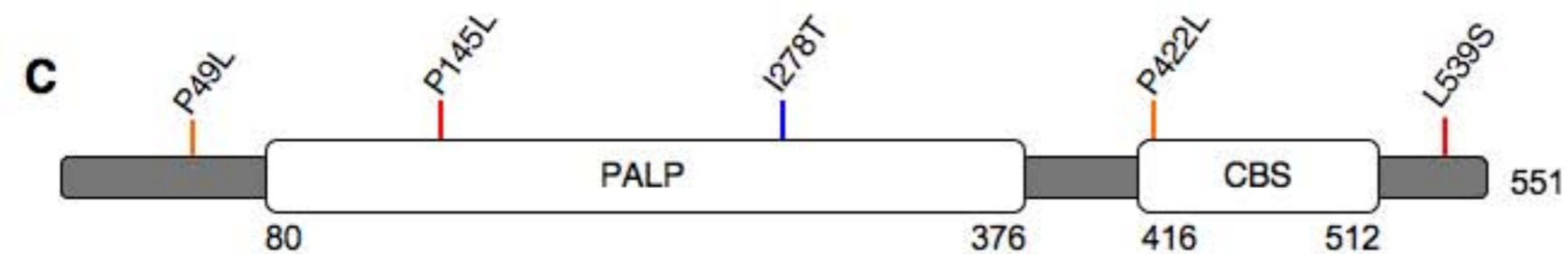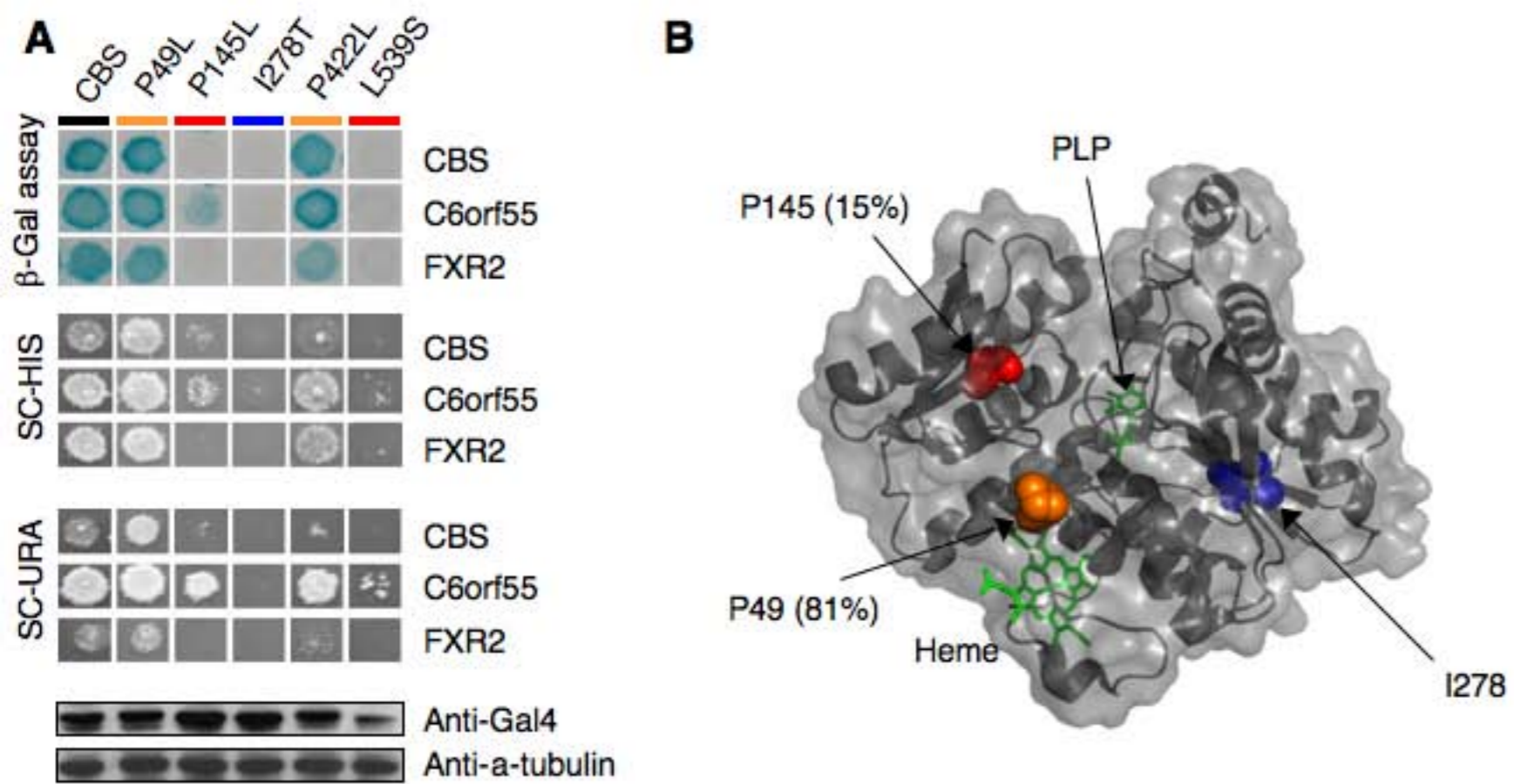
D

Y2H Reporter	CBS	P49L	P145L	I278T	P422L	L539S	Interactors
lacZ	yes	yes	no	no	yes	no	CBS
	yes	yes	yes	no	yes	no	C6orf55
	yes	yes	no	no	yes	no	FXR2
HIS3	yes	yes	yes	no	yes	no	CBS
	yes	yes	yes	no	yes	yes	C6orf55
	yes	yes	no	no	yes	yes	FXR2
URA3	yes	yes	yes	no	yes	no	CBS
	yes	yes	yes	no	yes	yes	C6orf55
	yes	yes	no	no	yes	no	FXR2
Interaction score	wt	+	R	-	+	-	CBS
	wt	+	+	-	+	R	C6orf55
	wt	+	-	-	+	-	FXR2
Enzyme activity	100%	NA	0%	1~5%	100%	NA	

Figure S2

A

	HGD	L25P	E42A	R225H	V300G	H371R
β-Gal assay						

SC-HIS

SC-URA

HGD, NUDT18, NIF3L1

Anti-Gal4
Anti-a-tubulin

B

Fe²⁺
H371 (25%)
V300
R225 (81%)
L25 (17%)
E42 (21%)

C

L25P E42A   R225H   V300G   H371R

HgmA   445
3   434

D

Y2H Reporter	HGD	L25P	E42A	R225H	V300G	H371R	Interactors
lacZ	yes	yes	no	yes	no	yes	HGD
	yes	yes	no	no	no	yes	NUDT18
	yes	yes	no	no	no	yes	NIF3L1
HIS3	yes	yes	no	yes	no	yes	HGD
	yes	yes	no	no	no	yes	NUDT18
	yes	yes	no	no	no	yes	NIF3L1
URA3	yes	yes	no	no	no	yes	HGD
	no	no	no	no	no	no	NUDT18
	yes	no	no	no	no	yes	NIF3L1
Interaction score	wt	+	-	R	-	+	HGD
	wt	+	-	-	-	+	NUDT18
	wt	R	-	-	-	+	NIF3L1
Enzyme activity	100%	NA	29%	0.1%	1.9%	0%	

Figure S3

A

β-Gal assay / SC-HIS / SC-URA panels with columns ACTG1, T89I, K118M, P264L, T278I, P332A; rows ACTB, ACTG1, CFL1, CFL2, DSTN; Anti-Gal4, Anti-a-tubulin.

B

Structure with ATP, P264, T278, P332 (19%), K118 (33%), T89 labeled.

C

Actin diagram, residues 2–375, with T89I, K118M, P264L, T278I, P332A marked.

D

Y2H reporter	ACTG1	T89I	K118M	P264L	T278I	P332A	Interactors
LacZ	yes	yes	yes	no	yes	yes	ACTB
	yes	yes	yes	no	yes	yes	ACTG1
	yes	yes	yes	no	yes	no	CF1
	yes	yes	yes	no	yes	no	CF2
	yes	yes	yes	no	yes	no	DSTN
HIS3	yes	yes	yes	yes	yes	yes	ACTB
	yes	yes	yes	yes	yes	yes	ACTG1
	yes	yes	yes	yes	yes	yes	CF1
	yes	yes	yes	no	yes	yes	CF2
	yes	yes	yes	no	yes	yes	DSTN
URA3	yes	yes	yes	yes	yes	yes	ACTB
	yes	yes	yes	yes	yes	yes	ACTG1
	yes	yes	yes	no	yes	no	CF1
	yes	yes	yes	no	yes	no	CF2
	yes	yes	yes	no	yes	no	DSTN
Interaction score	wt	+	+	R	+	+	ACTB
	wt	+	+	R	+	+	ACTG1
	wt	+	+	-	+	-	CF1
	wt	+	+	-	+	-	CF2
	wt	+	+	-	+	-	DSTN

Figure S4

A

	CDK4	R24C	R24H	N41S	S52N	
β-Gal assay						CCND3
						CDKN2B
						CDKN2D
						CDKN2C
SC-HIS						CCND3
						CDKN2B
						CDKN2D
						CDKN2C
SC-URA						CCND3
						CDKN2B
						CDKN2D
						CDKN2C
						Anti-Gal4
						Anti-a-tubulin

B

R24 (50%)
N41 (23%)
D67
D76
S52 (42%)

C

R24C
R24H
N41S
S52N

PKinase

6   295

303

D

Y2H reporter	CDK4	R24C	R24H	N41S	S52N	Interactors
LacZ	yes	yes	yes	yes	yes	CCND3
	yes	yes	yes	yes	yes	CDKN2B
	yes	yes	yes	yes	yes	CDKN2D
	yes	yes	yes	yes	yes	CDKN2C
HIS3	yes	yes	yes	yes	yes	CCND3
	yes	yes	yes	yes	yes	CDKN2B
	yes	yes	yes	yes	yes	CDKN2D
	yes	yes	yes	yes	yes	CDKN2C
URA3	yes	yes	yes	yes	yes	CCND3
	yes	yes	no	yes	yes	CDKN2B
	yes	yes	yes	yes	yes	CDKN2D
	yes	no	no	yes	yes	CDKN2C
Interaction score	wt	+	+	+	+	CCND3
	wt	+	R	+	+	CDKN2B
	wt	+	+	+	+	CDKN2D
	wt	R	R	+	+	CDKN2C

Figure S5

A

β-Gal assay / SC-HIS / SC-URA panels; columns: PRKAR1A, Q28X, Q37X, R42X, K63X, R74C, R97X, Q166X, R228X, Q304X, S307X; rows: AKAP10, MGC13057, PLEKHF2. Anti-Gal4 and Anti-α-tubulin blots.
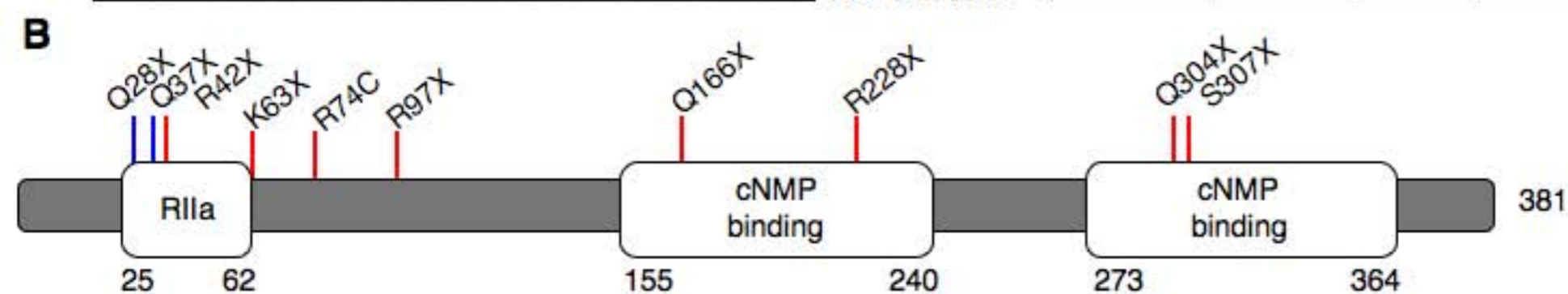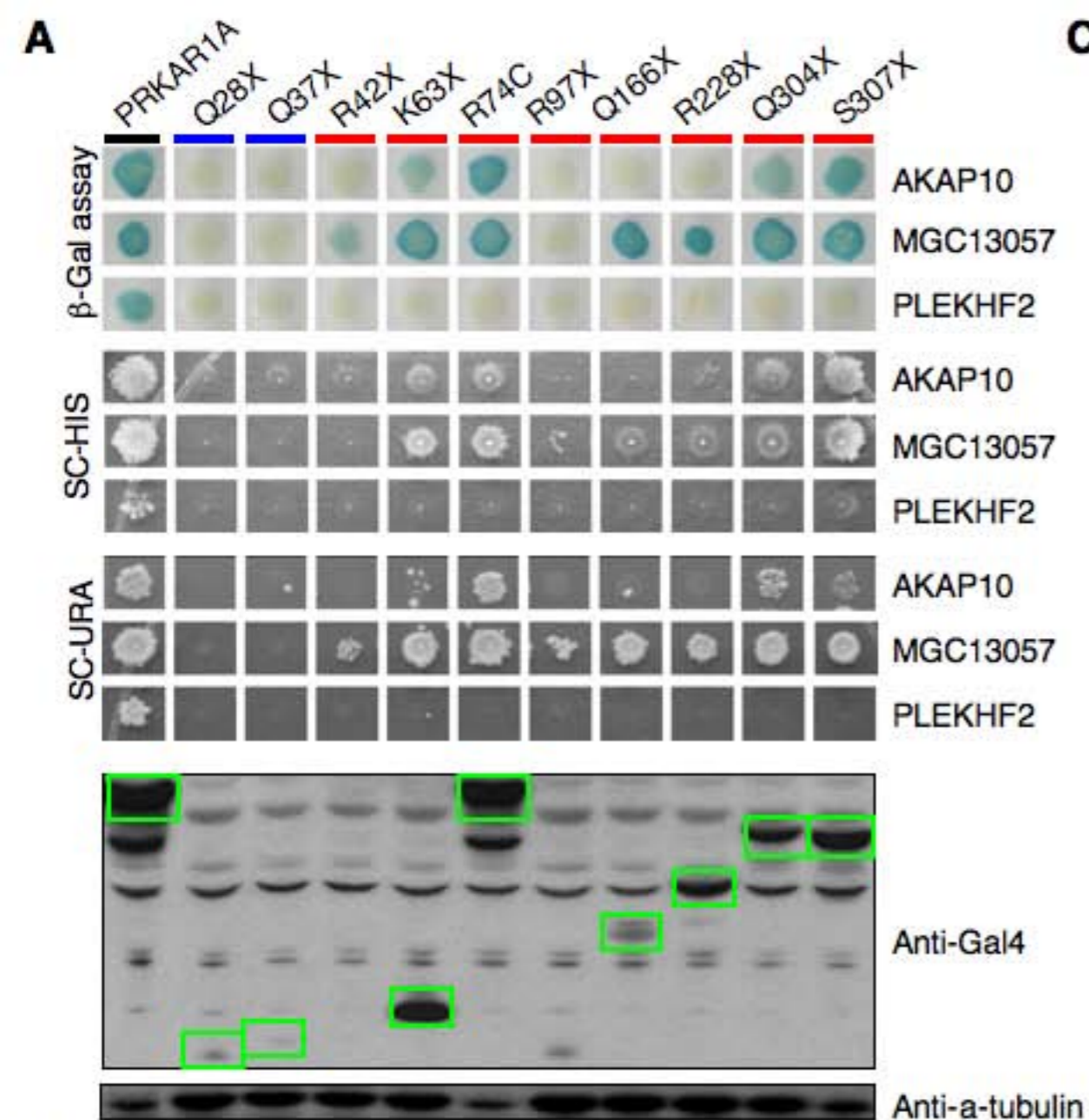
C

Y2H reporter	PRKAR1A	Q28X	Q37X	R42X	K63X	R74C	Interactors
lacZ	yes	no	no	no	yes	yes	AKAP10
	yes	no	no	yes	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
HIS3	yes	yes	yes	yes	yes	yes	AKAP10
	yes	no	no	no	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
URA3	yes	no	no	no	yes	yes	AKAP10
	yes	no	no	yes	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
Interaction score	wt	-	-	-	+	+	AKAP10
	wt	-	-	R	+	+	MGC13057
	wt	-	-	-	-	-	PLEKHF2

Y2H reporter	PRKAR1A	R97X	Q166X	R228X	Q304X	S307X	Interactors
lacZ	yes	no	no	no	yes	yes	AKAP10
	yes	no	yes	yes	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
HIS3	yes	no	no	yes	yes	yes	AKAP10
	yes	yes	yes	yes	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
URA3	yes	no	no	no	yes	yes	AKAP10
	yes	yes	yes	yes	yes	yes	MGC13057
	yes	no	no	no	no	no	PLEKHF2
Interaction score	wt	-	-	-	+	+	AKAP10
	wt	R	+	+	+	+	MGC13057
	wt	-	-	-	-	-	PLEKHF2

B

Protein domain diagram of PRKAR1A (381 aa) with RIIa domain (25–62), cNMP binding domain (155–240), and cNMP binding domain (273–364). Mutation positions marked: Q28X, Q37X, R42X, K63X, R74C, R97X, Q166X, R228X, Q304X, S307X.

Figure S6

Fig. S7

Fig. S8

Y2H reporter	Control 1	Control 2	Control 3	Control 4	Control 5
lacZ	no	yes	yes	yes	yes
HIS3	no	yes	yes	yes	yes
URA3	no	no	yes	yes	yes
Interaction score	-	+	+	+	+

Fig. S9

Supplemental table 1. The number of traits, genes, diseases and total mutations in each bin in Figure 2C and Figure S1 for the analysis of "in-frame" versus "truncating" mutations

Minimal fraction of "in-frame mutations in each bin			0	0.2	0.4	0.6	0.8	1
Figure 2C	Autosomal dominant	Traits	411	357	305	240	177	95
		Genes	329	288	249	198	151	83
		Diseases	363	315	268	213	164	90
		Total mutations	17085	13470	9258	5738	3500	1153
	Autosomal recessive	Traits	515	443	355	240	76	12
		Genes	482	414	335	230	74	12
		Diseases	469	404	330	227	72	12
		Total mutations	16539	14967	13120	8531	1525	101
Figure S1	Autosomal dominant non-essential	Traits	242	211	191	157	116	65
		Genes	203	177	162	135	103	60
		Diseases	226	197	177	149	115	65
		Total mutations	7587	7167	5871	3770	2371	812
	Autosomal recessive non-essential	Traits	421	362	290	200	65	10
		Genes	398	341	277	194	64	10
		Diseases	388	333	271	191	62	10
		Total mutations	12356	11101	9616	7107	1318	89

Supplemental table 2. Different mutation spectrums of different diseases associated with the same gene

Disease gene	Disease 1		Disease 2		Proprotion of in-frame mutations		P value
	OMIM ID	OMIM disease	OMIM ID	OMIM disease	Disease 1	Disease 2	
ABCA12	242500	ICHTHYOSIS CONGENITA, HARLEQUIN FETUS TYPE	601277	LAMELLAR ICHTHYOSIS 2	0.17	1.00	3.39X10^{-3}
ABCA4	601718	RETINITIS PIGMENTOSA 19	153800	AGE-RELATED MACULAR DEGENERATION 2	0.14	0.79	3.64X10^{-3}
ABCA4	601718	RETINITIS PIGMENTOSA 19	248200	STARGARDT DISEASE 1	0.14	0.63	1.28X10^{-2}
APOA1	604091	PRIMARY HYPOALPHALIPOPROTEINEMIA	105200	FAMILIAL VISCERAL AMYLOIDOSIS	0.42	1.00	4.45X10^{-3}
APOB	107730	FAMILIAL HYPOBETALIPOPROTEINEMIA 1	144010	HYPERCHOLESTEROLEMIA, AUTOSOMAL DOMINANT TYPE B	0.21	1.00	2.50X10^{-5}
ARX	300215	LISSENCEPHALY WITH AMBIGUOUS GENITALIA	300419	MENTAL RETARDATION, X-LINKED 54	0.44	1.00	1.89X10^{-2}
ATM	208900	ATAXIA-TELANGIECTASIA	114480	BREAST CANCER	0.19	0.76	3.94X10^{-9}
BRCA1	113705	BREAST-OVARIAN CANCER	176807	PROSTATE CANCER	0.31	1.00	2.96X10^{-3}
BRCA1	114480	BREAST CANCER	176807	PROSTATE CANCER	0.21	1.00	4.44X10^{-4}
BRCA1	114480	BREAST CANCER	113705	BREAST-OVARIAN CANCER	0.21	0.31	3.12X10^{-3}
BRCA2	114480	BREAST CANCER	260350	PANCREATIC CARCINOMA	0.21	0.70	1.54X10^{-3}
BRCA2	176807	PROSTATE CANCER	260350	PANCREATIC CARCINOMA	0.00	0.70	2.56X10^{-2}
BRCA2	600185	BREAST CANCER, TYPE 2	114480	BREAST CANCER	0.11	0.21	4.62X10^{-3}
BRCA2	600185	BREAST CANCER, TYPE 2	260350	PANCREATIC CARCINOMA	0.11	0.70	4.30X10^{-5}
BRCA2	604370	EPITHELIAL OVARIAN CANCER	260350	PANCREATIC CARCINOMA	0.07	0.70	3.48X10^{-4}
BRCA2	605724	FANCONI ANEMIA, COMPLEMENTATION GROUP D1	260350	PANCREATIC CARCINOMA	0.27	0.70	4.86X10^{-2}
CACNA1A	108500	EPISODIC ATAXIA, TYPE 2	141500	FAMILIAL HEMIPLEGIC MIGRAINE 1	0.34	0.91	1.33X10^{-3}
CDH1	137215	GASTRIC CANCER	176807	PROSTATE CANCER	0.30	0.80	4.11X10^{-2}
CDH23	601067	USHER SYNDROME, TYPE ID	601386	DEAFNESS, AUTOSOMAL RECESSIVE 12	0.32	1.00	1.91X10^{-7}
CFTR	219700	CYSTIC FIBROSIS	277180	CONGENITAL BILATERAL APLASIA OF VAS DEFERENS	0.49	0.83	2.59X10^{-9}
CFTR	219700	CYSTIC FIBROSIS	167800	PANCREATITIS, HEREDITARY	0.49	1.00	2.96X10^{-2}
COL1A1	166200	OSTEOGENESIS IMPERFECTA, TYPE I	166210	OSTEOGENESIS IMPERFECTA, TYPE IIA	0.19	0.93	2.80X10^{-21}
COL1A1	166200	OSTEOGENESIS IMPERFECTA, TYPE I	166220	OSTEOGENESIS IMPERFECTA, TYPE IV	0.19	0.75	2.92X10^{-6}
COL1A1	166200	OSTEOGENESIS IMPERFECTA, TYPE I	259420	OSTEOGENESIS IMPERFECTA, TYPE III	0.19	0.83	8.15X10^{-9}
COL1A1	166200	OSTEOGENESIS IMPERFECTA, TYPE I	120150	OI/EDS COMBINED SYNDROME	0.19	0.82	8.41X10^{-7}
COL1A1	166220	OSTEOGENESIS IMPERFECTA, TYPE IV	166210	OSTEOGENESIS IMPERFECTA, TYPE IIA	0.75	0.93	3.69X10^{-2}
COL1A2	130060	EHLERS-DANLOS SYNDROME, TYPE VII	166200	OSTEOGENESIS IMPERFECTA, TYPE I	0.08	0.80	1.66X10^{-4}
COL1A2	130060	EHLERS-DANLOS SYNDROME, TYPE VII	166210	OSTEOGENESIS IMPERFECTA, TYPE IIA	0.08	0.87	4.18X10^{-6}
COL1A2	130060	EHLERS-DANLOS SYNDROME, TYPE VII	166220	OSTEOGENESIS IMPERFECTA, TYPE IV	0.08	0.81	6.98X10^{-6}
COL1A2	130060	EHLERS-DANLOS SYNDROME, TYPE VII	259420	OSTEOGENESIS IMPERFECTA, TYPE III	0.08	0.95	7.14X10^{-7}
COL2A1	108300	STICKLER SYNDROME, TYPE I	151210	PLATYSPONDYLIC LETHAL SKELETAL DYSPLASIA, TORRANCE TYPE	0.10	0.56	4.96X10^{-3}
COL2A1	108300	STICKLER SYNDROME, TYPE I	184250	SPONDYLOEPIMETAPHYSEAL DYSPLASIA, STRUDWICK TYPE	0.10	1.00	8.21X10^{-5}

Gene	OMIM	Disease 1	OMIM	Disease 2			
COL2A1	108300	STICKLER SYNDROME, TYPE I	200610	ACHONDROGENESIS, TYPE II	0.10	1.00	1.15×10^{-12}
COL2A1	151210	PLATYSPONDYLIC LETHAL SKELETAL DYSPLASIA, TORRANCE TYPE	200610	ACHONDROGENESIS, TYPE II	0.56	1.00	5.31×10^{-3}
COL2A1	156550	KNIEST DYSPLASIA	184250	SPONDYLOEPIMETAPHYSEAL DYSPLASIA, STRUDWICK TYPE	0.33	1.00	2.94×10^{-2}
COL2A1	156550	KNIEST DYSPLASIA	200610	ACHONDROGENESIS, TYPE II	0.33	1.00	4.71×10^{-5}
COL6A3	254090	ULLRICH CONGENITAL MUSCULAR DYSTROPHY	158810	BETHLEM MYOPATHY	0.38	0.91	4.08×10^{-2}
DMD	300376	MUSCULAR DYSTROPHY, BECKER TYPE	302045	DILATED CARDIOMYOPATHY, 3B	0.13	0.50	3.31×10^{-2}
DMD	310200	MUSCULAR DYSTROPHY, DUCHENNE TYPE	300376	MUSCULAR DYSTROPHY, BECKER TYPE	0.04	0.13	3.89×10^{-2}
DMD	310200	MUSCULAR DYSTROPHY, DUCHENNE TYPE	302045	DILATED CARDIOMYOPATHY, 3B	0.04	0.50	3.50×10^{-4}
F5	227400	FACTOR V DEFICIENCY	188055	THROMBOPHILIA DUE TO DEFICIENCY OF ACTIVATED PROTEIN C COFACTOR	0.30	0.86	7.93×10^{-3}
FGFR2	101600	PFEIFFER SYNDROME	123500	CROUZON SYNDROME	0.62	0.94	2.45×10^{-3}
FGG	202400	CONGENITAL AFIBRINOGENEMIA	134820	DYSFIBRINOGENEMIA CAUSING RECURRENT THROMBOSIS	0.62	0.96	6.74×10^{-3}
FLNA	300049	PERIVENTRICULAR HETEROTOPIA	304120	OTOPALATODIGITAL SYNDROME, TYPE II	0.29	1.00	1.31×10^{-3}
FLNB	272460	SPONDYLOCARPOTARSAL SYNOSTOSIS SYNDROME	150250	LARSEN SYNDROME, AUTOSOMAL DOMINANT	0.00	1.00	7.94×10^{-3}
GBA	230800	GAUCHER DISEASE, TYPE I	230900	GAUCHER DISEASE, TYPE II	0.75	0.94	3.07×10^{-2}
GBA	231000	GAUCHER DISEASE, TYPE III	230900	GAUCHER DISEASE, TYPE II	0.63	0.94	3.98×10^{-2}
GCK	125851	MATURITY-ONSET DIABETES OF THE YOUNG, TYPE II	602485	FAMILIAL HYPERINSULINEMIC HYPOGLYCEMIA 3	0.44	1.00	4.57×10^{-2}
GCK	125851	MATURITY-ONSET DIABETES OF THE YOUNG, TYPE II	606176	PERMANENT NEONATAL DIABETES MELLITUS	0.44	0.71	3.20×10^{-2}
GJB2	220290	NEUROSENSORY DEAFNESS, AUTOSOMAL RECESSIVE 1	148210	KERATITIS-ICHTHYOSIS-DEAFNESS SYNDROME	0.53	1.00	3.36×10^{-2}
HEXA	272800	TAY-SACHS DISEASE	230710	GANGLIOSIDOSIS, GM2, JUVENILE, A(M)B VARIANT	0.50	0.90	1.98×10^{-2}
IKBKG	308300	INCONTINENTIA PIGMENTI	300291	HYPOHIDROTIC ECTODERMAL DYSPLASIA WITH IMMUNE DEFICIENCY	0.13	0.69	2.49×10^{-4}
KCNQ1	220400	JERVELL AND LANGE-NIELSEN SYNDROME	192500	LONG QT SYNDROME 1	0.38	0.84	5.50×10^{-4}
KRT14	131900	EPIDERMOLYSIS BULLOSA SIMPLEX, KOEBNER TYPE	131760	EPIDERMOLYSIS BULLOSA HERPETIFORMIS, DOWLING-MEARA TYPE	0.61	1.00	1.05×10^{-2}
L1CAM	307000	HYDROCEPHALUS DUE TO CONGENITAL STENOSIS OF AQUEDUCT OF SYLVIUS	303350	MASA SYNDROME	0.32	0.64	1.93×10^{-3}
LAMA3	226700	JUNCTIONAL EPIDERMOLYSIS BULLOSA, HERLITZ TYPE	226650	JUNCTIONAL EPIDERMOLYSIS BULLOSA, NON-HERLITZ TYPE	0.00	0.44	1.72×10^{-2}
LHCGR	152790	GONADOTROPIN UNRESPONSIVENESS	176410	MALE-LIMITED PRECOCIOUS PUBERTY	0.67	1.00	2.82×10^{-2}
LRP5	259770	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME	133780	EXUDATIVE VITREORETINOPATHY 1	0.50	0.82	4.02×10^{-2}
MEN1	131100	MULTIPLE ENDOCRINE NEOPLASIA, TYPE I	145000	HYPERPARATHYROIDISM 1	0.30	0.60	3.29×10^{-3}
MLH1	609310	HEREDITARY NONPOLYPOSIS COLORECTAL CANCER, TYPE 2	114500	COLORECTAL CANCER	0.34	0.70	3.61×10^{-2}
MRPL36	609054	FANCONI ANEMIA, COMPLEMENTATION GROUP J	114480	BREAST CANCER	0.40	1.00	3.38×10^{-2}
MSH2	120435	LYNCH SYNDROME I	137215	GASTRIC CANCER	0.32	0.83	1.59×10^{-2}
MSH2	158320	MUIR-TORRE SYNDROME	114500	COLORECTAL CANCER	0.00	0.47	6.32×10^{-3}
MSH2	158320	MUIR-TORRE SYNDROME	137215	GASTRIC CANCER	0.00	0.83	3.87×10^{-4}

MSH2	158320	MUIR-TORRE SYNDROME	120435	LYNCH SYNDROME I	0.00	0.32	6.95×10^{-3}
MSH6	600678	HEREDITARY NONPOLYPOSIS COLORECTAL CANCER, TYPE 5	114500	COLORECTAL CANCER	0.25	0.54	2.54×10^{-2}
NF1	162200	NEUROFIBROMATOSIS, TYPE I	162210	FAMILIAL SPINAL NEUROFIBROMATOSIS	0.14	0.60	2.34×10^{-2}
NF1	162200	NEUROFIBROMATOSIS, TYPE I	601321	NEUROFIBROMATOSIS-NOONAN SYNDROME	0.14	0.57	1.03×10^{-2}
PAX6	106210	ANIRIDIA, TYPE II	165550	BILATERAL OPTIC NERVE HYPOPLASIA	0.22	0.88	2.49×10^{-4}
PMP22	118220	DEMYELINATING CHARCOT-MARIE-TOOTH DISEASE, TYPE 1A	145900	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS	0.58	0.95	8.36×10^{-3}
PMP22	162500	HEREDITARY NEUROPATHY WITH LIABILITY TO PRESSURE PALSIES	145900	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS	0.31	0.95	1.58×10^{-4}
POMC	609734	PROOPIOMELANOCORTIN DEFICIENCY	601665	OBESITY	0.00	0.80	6.99×10^{-3}
RAG1	601457	SEVERE COMBINED IMMUNODEFICIENCY, AUTOSOMAL RECESSIVE, T CELL-NEGATIVE, B CELL-NEGATIVE, NK CELL-POSITIVE	603554	OMENN SYNDROME	0.21	0.77	1.74×10^{-3}
RAG1	601457	SEVERE COMBINED IMMUNODEFICIENCY, AUTOSOMAL RECESSIVE, T CELL-NEGATIVE, B CELL-NEGATIVE, NK CELL-POSITIVE	179615	ALPHA/BETA T-CELL LYMPHOPENIA WITH GAMMA/DELTA T-CELL EXPANSION SEVERE CYTOMEGALOVIRUS INFECTION, AND AUTOIMMUNITY	0.21	1.00	1.03×10^{-3}
RET	142623	SUSCEPTIBILITY TO HIRSCHSPRUNG DISEASE 1	171400	MULTIPLE ENDOCRINE NEOPLASIA, TYPE IIA	0.65	1.00	3.85×10^{-5}
RET	142623	SUSCEPTIBILITY TO HIRSCHSPRUNG DISEASE 1	188550	PAPILLARY THYROID CARCINOMA	0.65	1.00	1.09×10^{-5}
ROR2	113000	BRACHYDACTYLY, TYPE B1	268310	ROBINOW SYNDROME	0.00	0.45	4.45×10^{-2}
SCN1A	607208	SEVERE MYOCLONIC EPILEPSY OF INFANCY	604233	GENERALIZED EPILEPSY WITH FEBRILE SEIZURES PLUS	0.51	1.00	1.08×10^{-3}
SCN5A	113900	PROGRESSIVE FAMILIAL HEART BLOCK, TYPE IA	603830	LONG QT SYNDROME 3	0.76	1.00	8.13×10^{-3}
SCN5A	601144	BRUGADA SYNDROME 1	603830	LONG QT SYNDROME 3	0.77	1.00	6.26×10^{-4}
SCN5A	603829	PAROXYSMAL FAMILIAL VENTRICULAR FIBRILLATION	603830	LONG QT SYNDROME 3	0.60	1.00	1.22×10^{-2}
SETX	606002	SPINOCEREBELLAR ATAXIA, AUTOSOMAL RECESSIVE 1	602433	JUVENILE AMYOTROPHIC LATERAL SCLEROSIS 4	0.33	1.00	3.25×10^{-2}
VHL	193300	VON HIPPEL-LINDAU SYNDROME;	171300	PHEOCHROMOCYTOMA	0.51	0.95	4.34×10^{-5}
WFS1	222300	WOLFRAM SYNDROME 1	600965	NONSYNDROMIC SENSORINEURAL DEAFNESS, AUTOSOMAL DOMINANT 6	0.60	1.00	6.00×10^{-4}
WT1	136680	FRASIER SYNDROME	194080	DENYS-DRASH SYNDROME	0.17	0.78	7.79×10^{-3}
WT1	136680	FRASIER SYNDROME	256370	EARLY-ONSET NEPHROTIC SYNDROME WITH DIFFUSE MESANGIAL SCLEROSIS	0.17	1.00	1.52×10^{-2}
WT1	194070	WILMS TUMOR 1	194080	DENYS-DRASH SYNDROME	0.20	0.78	9.17×10^{-6}
WT1	194070	WILMS TUMOR 1	256370	EARLY-ONSET NEPHROTIC SYNDROME WITH DIFFUSE MESANGIAL SCLEROSIS	0.20	1.00	4.74×10^{-4}

Supplementary table 3. Enrichment of in-frame disease mutations in different Pfam domains causing different disease

Gene symbol	OMIM ID	OMIM disease	Mode of Inheritance	Pfam domain (residues)	Fold enrichment	P value
COMP	132400	MULTIPLE EPIPHYSEAL DYSPLASIA 1	autosomal dominant	TSP 3 (1080-1116)	16.3	1.33×10^{-4}
	132400	MULTIPLE EPIPHYSEAL DYSPLASIA 1	autosomal dominant	TSP 3 (1188-1224)	9.5	2.50×10^{-2}
	177170	PSEUDOACHONDROPLASIA	autosomal dominant	TSP 3 (1302-1347)	5.9	1.99×10^{-2}
	177170	PSEUDOACHONDROPLASIA	autosomal dominant	TSP 3 (1371-1407)	6.1	4.32×10^{-2}
	177170	PSEUDOACHONDROPLASIA	autosomal dominant	TSP 3 (1410-1455)	13.5	2.35×10^{-7}
	177170	PSEUDOACHONDROPLASIA	autosomal dominant	TSP 3 (1518-1563)	13.5	2.35×10^{-7}
DMD	300376	MUSCULAR DYSTROPHY, BECKER TYPE	unknown	CH (405-720)	51.6	5.27×10^{-3}
	310200	MUSCULAR DYSTROPHY, DUCHENNE TYPE	x-linked recessive	ZZ (9921-10056)	22.5	1.41×10^{-2}
FLNA	300049	PERIVENTRICULAR HETEROTOPIA	x-linked dominant	CH (132-447)	18.3	4.99×10^{-2}
	304120	OTOPALATODIGITAL SYNDROME, TYPE II	x-linked dominant	CH (501-807)	33.7	1.79×10^{-3}
	309350	MELNICK-NEEDLES SYNDROME	x-linked dominant	Filamin (3477-3738)	Infinite	9.12×10^{-4}
	311300	OTOPALATODIGITAL SYNDROME, TYPE I	x-linked dominant	CH (501-807)	Infinite	5.62×10^{-5}
PTPN11	151100	LEOPARD SYNDROME 1	autosomal dominant	Y phosphatase (819-1560)	Infinite	2.62×10^{-3}
	163950	NOONAN SYNDROME 1	autosomal dominant	SH2 (18-243)	5.2	8.17×10^{-6}
RYR1	117000	CENTRAL CORE DISEASE OF MUSCLE	autosomal dominant	Iontrans (14421-14811)	37.8	2.41×10^{-20}
	145600	SUSCEPTIBILITY TO MALIGNANT HYPERTHERMIA 1	autosomal dominant	RYDR ITPR (6471-7095)	7.5	4.28×10^{-8}
SCN5A	601144	BRUGADA SYNDROME 1	autosomal dominant	Iontrans (477-1236)	2.6	1.77×10^{-2}
	603830	LONG QT SYNDROME 3	autosomal dominant	Iontrans (4686-5313)	3.4	1.66×10^{-2}
SPTB	130600	ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE	autosomal dominant	Spectrin (6027-6210)	Infinite	9.46×10^{-9}
	182870	SPHEROCYTOSIS, TYPE I	autosomal dominant	CH (522-834)	19.7	3.88×10^{-2}
TP73L	106260	AEC SYNDROME	autosomal dominant	SAM 2 (1506-1704)	Infinite	3.21×10^{-10}
	129400	RAPP-HODGKIN SYNDROME	autosomal dominant	SAM 2 (1506-1704)	Infinite	3.22×10^{-3}
	604292	EEC SYNDROME 3	autosomal dominant	P53 (372-960)	Infinite	3.52×10^{-11}
WAS	300299	SEVERE CONGENITAL NEUTROPENIA, X-LINKED	x-linked recessive	PBD (711-888)	Infinite	4.11×10^{-2}
	313900	THROMBOCYTOPENIA 1	x-linked recessive	WH1 (117-435)	10.8	6.40×10^{-23}

Supplementary table 4. Primers used for cloning disease-causing mutant alleles

Gene	Primer Type	Primer Name	Sequence
CBS	attB1.1 forward	CBS-F01	GGGGACAACTTTGTACAAAAAAGTTGGCatgccttctgagaccccccag
	attB2.1 reverse	CBS-R01	GGGGACAACTTTGTACAAGAAAGTTGacttctggtcccgctcctg
	forward internal	CBS-c146t-F	GGATCCGGCCCGATGCTCTGAGCAGGTGCACCTGGC
		CBS-c434t-F	CCCGGGGACACGATTATCGAGCTGACATCCGGGAACACCGGG
		CBS-t833c-F	AGAAGTGTCCTGGATGCAGGATCACTGGGGTGGATCCCGAAGGG
		CBS-c1265t-F	AGGAGCTGGGCCTGTCAGCCCTGCTGACCGTGCTCCCGACCA
		CBS-t1616c-F	CGGGGTGGTCACCGCCATTGACTCGCTGAACTTCGTGGCCG
	reverse internal	CBS-c146t-RM03	TGCCAGGTGCACCTGCTCAGAGCATCGGGCCGGATCC
		CBS-c434t-RM06	CCGGTGTTCCCGGATGTCAGCTCGATAATCGTGTCCCCG
		CBS-t833c-RM09	TGGACCCTTCGGGATCCACCCCAGTGATCCTGCATCCAGGA
		CBS-c1265t-RM12	TGGTCGGGAGCACGGTCAGCAGGGCTGACAGGCCCAGCTCC
		CBS-t1616c-RM15	CTGGGCGGCCACGAAGTTCAGCGAGTCAATGGCGGTGACCACC
HGD	attB1.1 forward	HGD-F01	GGGGACAACTTTGTACAAAAAAGTTGGCatggctgagttaaagtacatttctgga
	attB2.1 reverse	HGD-R01	GGGGACAACTTTGTACAAGAAAGTTGaattaggttctgctgggttcctgg
	forward internal	HGD-t74c-F	AGGATCCTCGCTGCCCAGGTTCCCGCCAGAAGGACAGAATAATC
		HGD-a125c-F	GCCCCTACAATCTCTATGCTGCGCAGCTCTCAGGATCGGC
		HGD-g674a-F	CAATGGCTTGGCCAATCCTCATGATTTCTTGATACCCATTG
		HGD-t899g-F	CAGACCCATCCATTTTCACAGGATTGACTGCTAAGTCTG
		HGD-a1112g-F	ACAGCACAATGACCCCCCCGTGGACCTGATGCTGACTGCT
	reverse internal	HGD-t74c-R	TTATTCTGTCCTTCTGGCGGGGAACCTGGGCAGCGAGGA
		HGD-a125c-R	GCCGATCCTGAGAGCTGCGCAGCATAGAGATTGTAGG
		HGD-g674a-R	ATGGGTATCAAGAAATCATGAGGATTGGCCAAGCCAT
		HGD-t899g-R	ACAGACTTAGCAGTCAATCCTGTGAAAATGGATGGGTCT
		HGD-a1112g-R	CAGTCAGCATCAGGTCCACGGGGGGGTCATTGTGCTG
ACTG1	attB1.1 forward	ACTG1-F01	GGGGACAACTTTGTACAAAAAAGTTGGCatggaagaagagatcgcc
	attB2.1 reverse	ACTG1-R01	GGGGACAACTTTGTACAAGAAAGTTGagaagcatttgcggtggacg
	forward internal	ACTG1-c266t-F	AGAAGATCTGGCACCACATCTTCTACAACGAGCTGCG
		ACTG1-a353t-F	CCAAGGCCAACAGAGAGATGATGACTCAGATTATGTTT
		ACTG1-c791t-F	CGGAGGCGCTGTTCCAGCTTTCCTTCCTGGGTATGGAAT
		ACTG1-c833t-F	GCGGCATCCACGAGACCATCTTCAACTCCATCATGAAGT
		ACTG1-c994g-F	AAGATCAAGATCATCGCAGCCCCAGAGCGCAAGTACTCG
	reverse internal	ACTG1-c266t-R	CGCAGCTCGTTGTAGAAGATGTGGTGCCAGATCTTCTC
		ACTG1-a353t-R	AACATAATCTGAGTCATCATCTCTCTGTTGGCCTTG
		ACTG1-c791t-R	TCCATACCCAGGAAGGAAAGCTGGAACAGCGCCTCCG
		ACTG1-c833t-R	TTCATGATGGAGTTGAAGATGGTCTCGTGGATGCCG
		ACTG1-c994g-R	AGTACTTGCGCTCTGGGGCTGCGATGATCTTGATCTT
CDK4	attB1.1 forward	CDK4-F01	GGGGACAACTTTGTACAAAAAAGTTGGCatggctacctctcgatatg
	attB2.1 reverse	CDK4-R01	GGGGACAACTTTGTACAAGAAAGTTGactccggattaccttcatccttat
	forward internal	CDK4-c70t-F	GGGACAGTGTACAAGGCCTGTGATCCCCACAGTGGC
		CDK4-g71a-F	GGACAGTGTACAAGGCCCATGATCCCCACAGTGGCC
		CDK4-a122g-F	AGAGTGTGAGAGTCCCCAGTGGAGGAGGAGGTGGAG
		CDK4-g155a-F	GAGGAGGCCTTCCCATCAACACAGTTCGTGAGGTGGCTT
	reverse internal	CDK4-c70t-R	GGCCACTGTGGGGATCACAGGCCTTGTACACTGTCCCAT
		CDK4-g71a-R	TGGCCACTGTGGGGATCATGGGCCTTGTACACTGTC
		CDK4-a122g-R	CCTCCACCTCCTCCTCCACTGGGGACTCTCACACTCTT
		CDK4-g155a-R	GCCACCTCACGAACTGTGTTGATGGGAAGGCCTCCTC
PRKAR1A	attB1.1 forward	PRKAR1A-F01	GGGGACAACTTTGTACAAAAAAGTTGGCatggagtctggcagtacc
	attB2.1 reverse	PRKAR1A-R01	GGGGACAACTTTGTACAAGAAAGTTGagacagacagtgacacaaac
	forward internal	PRKAR1A-c82t-F	GTCCAGAAGCATAACATTTAAGCGCTGCTCAAAGAT
		PRKAR1A-c109t-F	CTCAAAGATTCTATTGTGTAGTTGTGCACTGCTCGACC
		PRKAR1A-c124t-F	GTGCAGTTGTGCACTGCTTGACCTGAGAGACCCATGGC
		PRKAR1A-a187t-F	TTGGAGAAGGAGGAGGCATAACAGATTCAGAATCTGC
		PRKAR1A-c220t-F	CTGCAGAAAGCAGGCACTTGTACAGACTCAAGGGAGGAT
		PRKAR1A-c289t-F	GTTAAAGGTAGGAGGCGATGAGGTGCTATCAGCGCTGAG
		PRKAR1A-c496t-F	GCAGGAGAGACTGTGATTTAGCAAGGTGATGAAGGG
		PRKAR1A-c682t-F	AAATTGTGGGGCATCGACTGAGACAGCTATAGAAGAA
		PRKAR1A-c910t-F	GGGTCAGCTGCTGTGCTATAACGTCGGTCAGAAAATGAA
		PRKAR1A-c920g-F	CTGTGCTACAACGTCGGTGAGAAAATGAAGAGTTTGTT
	reverse internal	PRKAR1A-c82t-R	AATCTTTGAGCAGCGCTTAAATGTTATGCTTCTGGA
		PRKAR1A-c109t-R	GTCGAGCAGTGCACAACTACACAATAGAATCTTTGA
		PRKAR1A-c124t-R	CCATGGGTCTCTCAGGTCAAGCAGTGCACAACTGCACAA
		PRKAR1A-a187t-R	GCAGATTCTGAATCTGTTATGCCTCCTCCTTCTCCAACC
		PRKAR1A-c220t-R	CCTCCCTTGAGTCTGTACAAGTGCCTGCTTTCTGCAGAT
		PRKAR1A-c289t-R	CAGCGCTGATAGCACCTCATCGCCTCCTACCTTTAACC
		PRKAR1A-c496t-R	CCCCTTCATCACCTTGCTAAATCACAGTCTCTCCTGCGA
		PRKAR1A-c682t-R	TTCTTCTATAGCTGTCTCAGTCGATGCCCCACAATTT
		PRKAR1A-c910t-R	CATTTTCTGACCGACGTTATAGCACAGCAGCTGACC
		PRKAR1A-c920g-R	ACAAACTCTTCATTTTCTCACCGACGTTGTAGCACAGCA