TITLE

Evolution and Engineering of Allosteric Regulation in Protein Kinases

ONE SENTENCE SUMMARY

Cell signaling is easily rewired by introducing new phosphoregulation at latent allosteric surface sites.

AUTHORS

David Pincus^{1,#,*}, Jai P. Pandey^{1,#}, Pau Creixell^{2,3,4}, Orna Resnekov⁵, Kimberly A. Reynolds^{6,7,#,*}

equal contribution

* correspondence: pincus@wi.mit.edu; kimberly.reynolds@utsouthwestern.edu

AFFILIATIONS

¹Whitehead Institute for Biomedical Research, Cambridge, USA

²David H. Koch Institute for Integrative Cancer Research at MIT and ³Department of Biology and

⁴Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA

⁵1255 La Canada Road, Hillsborough, USA

⁶Green Center for Systems Biology and ⁷Department of Biophysics, University of Texas

Southwestern Medical Center, Dallas, USA

ABSTRACT

Allosteric regulation – the control of protein function by sites far from the active site – is a common feature of proteins that enables dynamic cellular responses. Reversible modifications such as phosphorylation are well suited to mediate such regulatory dynamics, yet the evolution of new allosteric regulation demands explanation. To understand this, we mutationally scanned the surface of a prototypical kinase to identify readily evolvable phosphorylation sites. The data reveal a set of spatially distributed “hotspots” that coevolve with the active site and preferentially modulate kinase activity. By engineering simple consensus phosphorylation sites at these hotspots we successfully rewired *in vivo* cell signaling. Beyond synthetic biology, the hotspots are frequently used by the diversity of natural allosteric regulatory mechanisms in the kinase family and exploited in human disease.

MAIN TEXT

Allosteric regulation requires the cooperative action of many amino acids to functionally link distantly positioned amino acids. As a consequence, it is difficult to understand how allostery can evolve through a process of stepwise variation and selection. However, members of a protein family often display diverse regulatory mechanisms, suggesting that despite the complex intramolecular cooperativity required, allostery evolves readily (1). A potential explanation for how this might occur comes from separate lines of work that indicate a latent capacity for regulation at a diversity of surfaces in proteins. For example, it is possible to engineer synthetic allosteric switches through domain insertion at certain surface sites (2-6), and screens for small molecules that modify protein function sometimes identify cryptic allosteric regulatory sites (7, 8). In addition, experimental analysis of regulation in orthologs of the yeast MAP kinase Fus3 indicates that the capacity for allosteric regulation existed well before the regulatory mechanism evolved (9). Taken together, these findings suggest that proteins have an internal architecture in which a few sites on the protein surface are functionally “pre-wired” to provide control of protein active sites (10). This pre-wiring has been proposed to result not as a consequence of the need for regulation, but simply from the need for proteins to be evolvable (11). Thus, the acquisition of new regulation might amount to engaging or activating preexisting allosteric networks, a route to the evolution of regulation that is consistent with stepwise variation and selection.

An excellent model to test this proposal is the eukaryotic protein kinases (EPKs), a protein family that has diversified to control a vast array of cellular signaling activities. The EPKs catalyze the transfer of a phosphate group from adenosine triphosphate (ATP) onto a Ser/Thr/Tyr residue of a substrate protein, a reaction that is subject to regulation by different mechanisms at many surface regions in members of the kinase family (Fig. 1), including protein-protein interactions, auto-inhibition, dimerization, and post-translational modification (12). Recently, Ferrell and colleagues proposed an idea for the evolution of one such mechanism –

phosphoregulation – in which phosphorylation of a Ser/Thr/Tyr surface residue regulates protein activity. Since phosphorylation introduces a negative charge, the idea is that phosphoregulation might evolve simply by mutating an allosterically pre-coupled negatively charged residue (Asp/Glu) to a phosphorylatable residue (Ser/Thr/Tyr) (13). Thus, a constitutive negative charge at a latent allosteric site can be transformed into a regulated negative charge in a potentially stepwise manner (14).

To experimentally test this idea, we used the prototypical yeast CMGC kinase Kss1 as a model (see Supplementary Text for extended Kss1 background). Kss1 is a homolog of human ERK and is involved in signal transduction pathways that regulate yeast filamentous growth and the mating response (15-18). Kss1 activity can be quantitatively monitored in living yeast cells by its ability to specifically activate fluorescent transcriptional reporters of the mating pheromone response in the absence of its paralog, Fus3 (Fig. 2A). We conducted an unbiased alanine scan of all 40 Asp/Glu residues on the surface of Kss1 to determine which positions are functionally coupled to kinase activity. We integrated the resulting 40 Kss1 mutants as the only copy of Kss1 in the yeast genome, tagged at their C-terminus with a 3xFLAG epitope (Supplementary Tables 1 and 2). To test their activity, we assayed for induction of the pheromone-responsive *AGA1pr*-YFP reporter at four concentrations of the alpha factor mating pheromone (α F) by flow cytometry. Though all mutants maintained wild type-like expression levels, nine mutations altered *in vivo* kinase activity (Fig. 2B, C, Supplementary Fig. 1A). Three of these positions were identified as Kss1 mutants with a functional effect in previous studies (D117, D156, D321, Supplementary Table 3) (19, 20). Though enriched in the N-terminal half of the primary Kss1 sequence, these nine mutations occur at positions distributed broadly over the Kss1 atomic structure - consistent with the notion that certain surface sites are selectively pre-wired to allosterically influence active site function.

We next tested whether these positions can support new regulation of Kss1 through phosphorylation by another yeast kinase *in vivo*. In principle, this gain of function can effectively rewire signaling through the mating response pathway. We chose to engineer regulation of Kss1 by protein kinase A (PKA) because the PKA substrate consensus motif, RRxS/T requires minimal local modifications, its activity in yeast cells is orthogonal to the pheromone pathway, and it can be hyper-activated in yeast via ectopic expression of Ras2^{G19V} (Fig. 3A, Supplementary Fig. 1B, C). We selected three of the nine mutationally sensitive positions (D8, E68 and E70) with the highest PKA substrate scores predicted by the computational tool pkaPS (21). To claim PKA-mediated allosteric regulation of Kss1, we must demonstrate: 1) that Kss1 retains functionality following introduction of a local PKA consensus motif (RRx**D/E**, termed pka-D/E); 2) that Kss1 loses activity when the charge is neutralized (RRx**A**, termed pka-A); and 3) that Kss1 now displays PKA-dependent activity *in vivo* with introduction of a phosphorylatable residue (RRx**S**, termed pka-S) (Supplementary Fig. 1d). In this manner, a functional surface negative charged residue can neutrally acquire a substrate consensus sequence for a kinase and become a phosphoregulatory site with one step of variation.

Introduction of pka-E at position 68 resulted in Kss1 loss-of-function (Supplementary Fig. 1E, F), indicating that in this instance, the mutation of positions 65-66 to arginine to introduce the PKA site was not neutral. However, introducing the PKA consensus motif at positions D8 and E70 showed the complete expected pattern of activity for gain of phosphoregulation (Fig. 3B). For both sites, introduction of the two arginine residues upstream was near neutral, mutation of the negatively charged residue caused loss of function, and Kss1 pka-S activity depended on enhanced PKA activity via estradiol-induced expression of Ras2^{G19V} (Fig. 3B). Immunoprecipitation of the 3xFLAG-tagged Kss1 mutants followed by Western blot analysis supports this finding. Both the pka-A and pka-S variants displayed activation loop phosphorylation when treated with alpha factor, indicating that they remain substrates of Ste7.

However, only Kss1-pka-S8 and Kss1-pka-S70 were recognized by an antibody specific for phosphorylated PKA substrates when purified from cells treated with estradiol (Fig. 3C). Moreover, both Kss1-pka-S8 and Kss1-pka-S70 were able to induce the morphological response to pheromone – the mating projection known as the “shmoo” – in an α F- and PKA activity-dependent fashion (Fig. 3B). Thus, the transcriptional and physiological outputs of Kss1 can be rewired to depend on an orthogonal input by a stepwise process of introducing a phosphorylation site at latent allosteric surface sites.

What is special about the nine surface negatively charged amino acids that they are allosterically pre-wired to regulate the Kss1 active site? Is this functional coupling idiosyncratic to Kss1 or conserved in the kinase family? To address this, we used the Statistical Coupling Analysis (SCA) (22-24) to examine the correlated conservation (or coevolution) of amino acid positions in an alignment encompassing all EPK subfamilies and a focused alignment of the CMGC subfamily that includes the MAP kinases (Supplementary Fig. 2A,B). The basic result from SCA is the finding that protein families have internal networks of coevolving amino acids (called “sectors”) that tend to link protein active sites to distantly positioned allosteric surface sites (22, 25-28). Consistent with this, we identified a protein sector in the EPK family that forms a physically contiguous network of amino acids within the three-dimensional structure (Supplementary Fig. 2C-E, Supplementary Table 4). The sector is enriched for positions associated with kinase function: comparison to a deep mutational scan of human ERK2 (29) shows a clear, statistically significant association between the sector and sites associated with loss-of-function ($p = 2.3E-19$ by Fisher Exact Test, Supplementary Fig. 3, Supplementary Table 5). Further, the sector encapsulates several structural motifs well known to be associated with kinase activation including the α C-helix, the DFG motif and the catalytic and regulatory spines (Supplementary Fig. 2C-E, Supplementary Fig. 4) (30, 31). Thus, like for other proteins, analysis

of conserved coevolution of amino acids in EPKs provides a sparse, distributed model for the functionally relevant energetic connectivity of amino acids (24).

Previous work has proposed that the sector represents the physical mechanism that underlies the pre-wiring of surface sites that serve as hotspots for the emergence of new allosteric regulation (10, 32). Consistent with this, eight of the nine functionally coupled surface D/E residues in Kss1 are sector connected ($p = 0.0098$, Fisher Exact Test) (Fig. 4A, B), including the two that yield new PKA-dependent phosphoregulation. Thus, the gain of new regulatory function in Kss1 occurs at sites that are not idiosyncratic, but that interact with an allosteric network that coevolves in the entire kinase family. This result is robust to details of alignment construction and statistical cutoffs for determining sector positions (Supplementary Fig. 5, Supplementary Table 6). These data support a model that new regulation preferentially emerges in proteins at surface sites that are evolutionarily prewired in protein families.

If so, all natural kinases should follow the principle that functionally sensitive and physiologically relevant allosteric sites, regardless of mechanism, should be found with statistical preference at sector-connected surfaces. The sector-connected surfaces would then provide an explanation for the diversity of regulatory sites observed in extant kinases (Fig. 1). To investigate this, we constructed a curated database of mutations sampled across a diversity of kinases (those listed in Fig 1A, Supplementary Table 7). These mutations were selected because they were experimentally demonstrated to disrupt kinase regulation and/or function, and, in many cases, are also associated with disease. An analysis of mutations sampled across the kinase superfamily reveals a clear pattern: functional mutations cluster around the sector edges with strong statistical preference ($p = 0.00068$, Fisher Exact Test) (Fig. 4c, d, Supplementary Table 8). Thus, we conclude that the natural architecture of the protein kinases does indeed facilitate the evolution of regulatory diversity.

165

166 The central idea supported by our results is that proteins contain a conserved cooperative

167 mechanism that endows specific sites on the protein surface with a latent capacity for allostery.

168 As a consequence of this intrinsic cooperative architecture, allosteric regulation may emerge in

169 a variety of mechanistic forms at multiple, distinct locations in different family members (Fig.

170 4E). Our results demonstrate a general strategy for engineering new cell signaling pathways –

171 *in vivo* phospho-regulation can in principle be introduced into any soluble protein by targeting

172 negatively charged residues at sector-connected surfaces (33). Further, the sector provides a

173 context for interpreting kinase mutations involved in disease (Fig. 4C), and suggests possible

174 cryptic sites for the development of allosteric inhibitors (8). Overall, this model provides a path

175 for understanding how complex regulatory systems evolve, and suggests that sector edges

176 provide a substrate for generating variation in cellular signaling and communication.

REFERENCES

1. J. Kuriyan, D. Eisenberg, The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983-990 (2007).
2. O. Dagliyan *et al.*, Rational design of a ligand-controlled protein conformational switch. *Proc Natl Acad Sci U S A* **110**, 6800-6804 (2013).
3. O. Dagliyan *et al.*, Engineering extrinsic disorder to control protein activity in living cells. *Science* **354**, 1441-1444 (2016).
4. G. Guntas, M. Ostermeier, Creation of an allosteric enzyme by domain insertion. *J Mol Biol* **336**, 263-273 (2004).
5. D. C. Nadler, S. A. Morgan, A. Flamholz, K. E. Kortright, D. F. Savage, Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat Commun* **7**, 12266 (2016).
6. K. A. Reynolds, R. N. McLaughlin, R. Ranganathan, Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564-1575 (2011).
7. J. A. Hardy, J. Lam, J. T. Nguyen, T. O'Brien, J. A. Wells, Discovery of an allosteric site in the caspases. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12461-12466 (2004).
8. P. Wu, M. H. Clausen, T. E. Nielsen, Allosteric small-molecule kinase inhibitors. *Pharmacol Ther* **156**, 59-68 (2015).
9. S. M. Coyle, J. Flores, W. A. Lim, Exploitation of latent allostery enables the evolution of new modes of MAP kinase regulation. *Cell* **154**, 875-887 (2013).
10. K. A. Reynolds, R. N. McLaughlin, R. Ranganathan, Hotspots for allosteric regulation on protein surfaces. *Cell* **147**, 1564-1575 (2011).
11. A. S. Raman, K. I. White, R. Ranganathan, Origins of Allostery and Evolvability in Proteins: A Case Study. *Cell* **166**, 468-480 (2016).

12. P. Pellicena, J. Kuriyan, Protein-protein interactions in the allosteric regulation of protein kinases. *Current opinion in structural biology* **16**, 702-709 (2006).
13. S. M. Pearlman, Z. Serber, J. E. Ferrell, Jr., A mechanism for the evolution of phosphorylation sites. *Cell* **147**, 934-946 (2011).
14. P. Creixell, E. M. Schoof, C. S. Tan, R. Linding, Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues. *Philos Trans R Soc Lond B Biol Sci* **367**, 2584-2593 (2012).
15. T. G. Boulton *et al.*, An insulin-stimulated protein kinase similar to yeast kinases involved in cell cycle control. *Science* **249**, 64-67 (1990).
16. W. E. Courchesne, R. Kunisawa, J. Thorner, A putative protein kinase overcomes pheromone-induced arrest of cell cycling in *S. cerevisiae*. *Cell* **58**, 1107-1119 (1989).
17. J. G. Cook, L. Bardwell, J. Thorner, Inhibitory and activating functions for MAPK Kss1 in the *S. cerevisiae* filamentous-growth signalling pathway. *Nature* **390**, 85-88 (1997).
18. O. Atay, J. M. Skotheim, Spatial and temporal signal processing and decision making by MAPK pathways. *J Cell Biol* **216**, 317-330 (2017).
19. A. B. Kusari, D. M. Molina, W. Sabbagh, Jr., C. S. Lau, L. Bardwell, A conserved protein interaction network involving the yeast MAP kinases Fus3 and Kss1. *The Journal of cell biology* **164**, 267-277 (2004).
20. H. D. Madhani, C. A. Styles, G. R. Fink, MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. *Cell* **91**, 673-684 (1997).
21. G. Neuberger, G. Schneider, F. Eisenhaber, pkaPS: prediction of protein kinase A phosphorylation sites with the simplified kinase-substrate binding model. *Biology direct* **2**, 1 (2007).
22. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774-786 (2009).

23. S. W. Lockless, R. Ranganathan, Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-299 (1999).
24. O. Rivoire, K. A. Reynolds, R. Ranganathan, Evolution-Based Functional Decomposition of Proteins. *PLoS Comput Biol* **12**, e1004817 (2016).
25. M. E. Hatley, S. W. Lockless, S. K. Gibson, A. G. Gilman, R. Ranganathan, Allosteric determinants in guanine nucleotide-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14445-14450 (2003).
26. A. I. Shulman, C. Larson, D. J. Mangelsdorf, R. Ranganathan, Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell* **116**, 417-429 (2004).
27. R. G. Smock *et al.*, An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Molecular systems biology* **6**, 414 (2010).
28. G. M. Suel, S. W. Lockless, M. A. Wall, R. Ranganathan, Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* **10**, 59-69 (2003).
29. L. Brenan *et al.*, Phenotypic Characterization of a Comprehensive Set of MAPK1/ERK2 Missense Mutants. *Cell Rep* **17**, 1171-1183 (2016).
30. A. P. Kornev, N. M. Haste, S. S. Taylor, L. F. Eyck, Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A* **103**, 17783-17788 (2006).
31. A. P. Kornev, S. S. Taylor, L. F. Ten Eyck, A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci U S A* **105**, 14377-14382 (2008).
32. J. Lee *et al.*, Surface sites for engineering allosteric control in proteins. *Science* **322**, 438-442 (2008).
33. D. Pincus, O. Resnekov, K. A. Reynolds, An evolution-based strategy for engineering allosteric regulation. *Phys Biol* **14**, 025002 (2017).

34. F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, T. Schwede, The SWISS-MODEL Repository and associated resources. *Nucleic acids research* **37**, D387-392 (2009).
35. M. S. Longtine *et al.*, Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953-961 (1998).
36. D. Pincus, K. Benjamin, I. Burbulis, A. E. Tsong, O. Resnekov, Reagents for investigating MAPK signalling in model yeast species. *Yeast* **27**, 423-430 (2010).
37. J. Pei, N. V. Grishin, PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in molecular biology* **1079**, 263-271 (2014).
38. O. Buzko, K. M. Shokat, A kinase sequence database: sequence alignments and family assignment. *Bioinformatics* **18**, 1274-1275 (2002).
39. M. F. Sanner, A. J. Olson, J. C. Spehner, Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305-320 (1996).
40. T. Tesileanu, L. J. Colwell, S. Leibler, Protein sectors: statistical coupling analysis versus conservation. *PLoS computational biology* **11**, e1004091 (2015).
41. D. Ma, J. G. Cook, J. Thorner, Phosphorylation and localization of Kss1, a MAP kinase of the *Saccharomyces cerevisiae* pheromone response pathway. *Mol Biol Cell* **6**, 889-909 (1995).
42. S. Pelet, Nuclear relocation of Kss1 contributes to the specificity of the mating response. *Sci Rep* **7**, 43636 (2017).
43. L. Bardwell, J. G. Cook, E. C. Chang, B. R. Cairns, J. Thorner, Signaling in the yeast pheromone response pathway: specific and high-affinity interaction of the mitogen-activated protein (MAP) kinases Kss1 and Fus3 with the upstream MAP kinase kinase Ste7. *Mol Cell Biol* **16**, 3637-3650 (1996).

ACKNOWLEDGEMENTS

This collaboration was initiated at the 2013 q-bio conference held at St. Johns College, Santa Fe, NM. We would like to thank R. Ranganathan for discussion and comments on the manuscript. We are grateful to the Whitehead Institute FACS facility and the Keck Microscopy facility for technical assistance. This work was supported by an NIH Early Independence Award (DP5 OD017941-01 to D.P.), the Green Center for Systems Biology, and the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative (Grant GBMF4557 to K.R.).

AUTHOR CONTRIBUTIONS

Conceptualization, K.A.R, D.P., and O.R.; Methodology, K.A.R., D.P., O.R. and P.C.; Investigation, D.P., J.P.P, O.R., and K.A.R., Writing – Original Draft, D.P. and K.A.R; Writing – Reviewing & Editing, D.P., J.P.P, O.R., and K.A.R, Supervision, D.P. and K.A.R.

FIGURE LEGENDS

Figure 1. Regulatory Diversity in the Eukaryotic Protein Kinases.

A. Unanchored dendrogram of the human kinome illustrating the diversity of the EPK superfamily and subfamilies. Individual subfamily members with functional mutations shown in Fig. 4c and included in Supplementary Table 7 are listed. TK: tyrosine kinase; TKL: TK-like; STE: STE7/11/20; CK1: Casein Kinase 1; AGC: protein kinase A/G/C; CAMK: Calmodulin kinase; CMGC: cyclin dependent kinase (CDK)/mitogen activated protein kinase (MAPK)/glycogen synthase kinase (GSK)/CDK-like kinase (CLK).

B. Allosteric regulatory sites from diverse kinases mapped to a single representative structure - yeast CDK Pho85 (PDB: 2PK9, shown as space-filled surface). Regulatory surfaces were identified by structural alignment of the kinase of interest to Pho85; all Pho85 positions within 4Å of the interaction surface are colored. Color coding is the same as in (A). This mapping shows that regulation occurs at structurally diverse sites across the kinase structure.

Figure 2. Alanine scan of acidic residues on the solvent accessible surface of yeast MAPK Kss1.

A. Schematic of the Kss1-dependent yeast pheromone pathway. The alpha factor (α F) mating pheromone binds to a G-protein coupled receptor (GPCR), leading to activation of a signaling cascade culminating at the MAPK Kss1. Kss1 then activates the Ste12 transcription factor to induce the mating transcriptional program, which can be monitored by fusing the promoter of the target gene *AGA1* to a YFP reporter.

B. Ribbon diagram of a Kss1 homology model (34) with the 40 solvent accessible Asp/Glu residues shown as spheres. The DFG motif and activation loop are indicated in light blue. All 40 positions were mutated individually to alanine to remove negative charge.

C. The 40 resulting yeast strains along with WT and *kss1* Δ controls were assayed for activation of the *AGA1pr*-YFP reporter by flow cytometry following treatment with 0, 0.01, 0.1 and 1 μ M α F

for 4 hours. Bars represent the average of the median YFP fluorescence from 3 biological replicates normalized to the untreated *kss1Δ* cells, and error bars are the standard deviation of the biological replicates. Mutations at red and green positions resulted in significantly reduced or increased YFP expression ($p < 0.05$) in response to at least two doses of α F, respectively. Yellow positions indicate that the mutation had no effect in this assay. The color coding is identical in (B). The data show that nine acidic positions on the solvent accessible surface are functionally coupled to kinase activity.

Figure 3. Engineering allosteric control of Kss1 by PKA phosphorylation.

A. Cartoon of the engineered PKA- and Kss1-dependent yeast pheromone pathway. In this schematic, Kss1 activation requires both activation loop phosphorylation by the upstream MAP2K Ste7, and phosphorylation by PKA at an allosterically coupled surface. To experimentally increase PKA activity, expression of constitutively activated Ras2(G19V) is induced by addition of estradiol, which in turn activates adenylate cyclase (AC) to generate cyclic AMP (cAMP) from ATP to activate PKA.

B. Kss1 mutants with PKA phosphorylation site consensus motifs introduced near position 8 (pka-X8, upper panel) or position 70 (pka-X70, lower panel) were assayed for expression of the *AGA1pr*-YFP reporter as in Fig. 2c. “X” stands for the amino acid at position 8 or 70 as denoted under the bar graphs. The images below the bar graphs show morphology and expression of the *AGA1pr*-YFP reporter in yeast cells bearing the indicated Kss1 mutants in the presence of 1 μ M alpha factor following growth in the presence or absence of 20 nM estradiol. The data indicate that phosphorylation by PKA at these positions can allosterically regulate Kss1 activity.

C. 3xFLAG-tagged wild type Kss1 and pka-X8 and -X70 mutants were immunoprecipitated from untreated cells or cells that had been treated with both 20 nM estradiol and 1 μ M alpha factor. IP eluates were analyzed by Western blotting for total Kss1 as well as Kss1 phosphorylated on its activation loop (phospho act. loop) or at the engineered PKA site (phospho pka site). Merged

images show that all mutants can be phosphorylated on their activation loop in the presence of alpha factor, but only pka-S8 and pka-S70 can be phosphorylated by PKA in the presence of estradiol.

Figure 4. Sector connected surface sites are hotspots for allosteric regulation.

A. Space filling diagram of a Kss1 homology model (34). The CMGC sector, defined as positions that co-evolve across the CMGC kinases, is indicated in blue. Acidic surface residues with a neutral, activating, or inactivating effect on kinase function upon mutation to alanine are shown as yellow, green or red spheres respectively.

B. Fisher's exact table demonstrating statistically significant enrichment of acidic surface residues with a functional effect upon mutation at sector-connected positions. To be sector connected, a position must have at least one atom within 4 Å of the sector.

C. The EPK superfamily-wide sector (blue spheres) mapped to the CMGC yeast kinase Pho85 (PDB: 2PK9, grey cartoon and surface). Red positions are sites collected from the literature known to alter kinase function when mutated in a functional study or human disease context (Supplementary table 7).

D. Fisher's exact table demonstrating statistically significant enrichment of the functional mutations shown in **c** at sector-connected positions.

E. Model for the evolution of regulatory diversity. Latent allosteric sites distributed across the protein surface (red circles) are connected to the active site via a protein sector (blue arrows). These sites are poised for the acquisition of new regulation via evolutionary, disease, or engineering processes. In any particular family member, only a subset of sites may be used, and the regulatory mechanism need not be conserved across homologs.

SUPPLEMENTARY MATERIALS

Supplementary Figures 1-6

Figure 1. Experimental approach to introduce a PKA phosphorylation site that controls

MAPK Kss1 activity

Figure 2. Statistical Coupling Analysis (SCA) of the Eukaryotic Protein Kinases (EPKs)

Figure 3. ERK2 mutations within the kinase sector are enriched for loss-of-function

Figure 4. The kinase sector encompasses the catalytic and regulatory spines

Figure 5. The relationship of negatively charged surface positions to the kinome-wide EPK

sector

Supplementary Tables 1-8

Table 1. Plasmids

Table 2. Yeast strains

Table 3. Comparison of Kss1 point mutations from the literature with our data

Table 4. List of sector positions for several representative kinases

Table 5. Statistical association between the sector, conservation and ERK2 mutational data

Table 6. Statistical association between the sector, conservation and KSS1 D/E surface

mutations

Table 7. Curated set of functional mutations for a diversity of kinases and references

Table 8. Statistical association between the sector, conservation and functional mutations

sampled across a diversity of kinases

Methods

Supplementary Text

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Experimental approach to introduce a PKA phosphorylation site that controls MAPK Kss1 activity.

A. Expression levels for all forty Kss1 mutants in which each acidic surface position was mutated to alanine, alongside WT and *kss1Δ* cell controls. Kss1 (and all mutants) were tagged with 3xFLAG and expression level was monitored by anti-FLAG immunoblot. Total protein loaded in each lane was monitored with an anti-Pgk1 antibody. Mutations are grouped into sector-connected and not sector-connected categories, as discussed later in the text. All alanine mutants show expression levels similar to WT.

B. To increase PKA activity, constitutively active Ras2(G19V) was expressed from a promoter activated in proportion to the concentration of estradiol in the media. Growth was monitored in log phase by measuring OD₆₀₀ over time in cells treated with the indicated concentrations of estradiol and plotted relative to cells without estradiol. Error bars are the standard deviation of three independent cultures. Based on these data, we chose to use 20 nM estradiol for all experiments because this is the highest concentration that did not result in growth inhibition.

C. Localization of YFP-Ras2 expressed from its endogenous promoter and YFP-Ras2(G19V) expressed at two concentrations of estradiol, showing significant over-expression at 20 nM but proper plasma membrane localization.

D. Schematic of the mutational strategy to introduce a functional PKA phosphorylation site at an allosterically coupled negatively charged surface position (red circle with “-” sign). First, two consecutive Arg residues are introduced by mutation (RRx) at the -2 and -3 positions with respect to the negatively charged (Asp/Glu) position to create the PKA consensus motif. While the Asp/Glu is maintained at position 0, the kinase must retain function in the presence of the RRx. Next position 0 is mutated to Ala in the context of the RRx to remove the negative charge. Removal of the negative charge should result in a loss-of-function kinase. Finally, mutation of position 0 to Ser must conditionally restore kinase activity in the presence of PKA activity.

E. Insertion of the RRx motif at position 68 resulted in Kss1 loss-of-function as assayed for activation of the *AGA1pr*-YFP reporter by flow cytometry following treatment with 0, 0.01, 0.1 and 1 μ M α F for 4 hours. Bars represent the average of the median YFP fluorescence from 3 biological replicates normalized to the untreated *kss1* Δ cells, and error bars are the standard deviation of the biological replicates.

F. Introduction of the RRx motif at position 68 resulted in reduced Kss1 protein expression and loss of activation loop phosphorylation. Samples of 3xFLAG-tagged WT and pka-E68 were monitored by anti-FLAG Western blot, and were also probed for phosphorylation of the activation loop (phospho act. loop)

Supplementary Figure 2. Statistical Coupling Analysis (SCA) of the Eukaryotic Protein

Kinases. The analysis was performed for two different multiple sequence alignments of the kinase catalytic domain: one specific to the CMGC kinases (635 sequences), and one containing 7128 kinases sampled across the kinome.

A. Histogram showing the distribution of pairwise sequence identities computed across all pairs of sequences in the CMGC alignment.

B. As in (A) but for the kinome wide alignment. Both alignments show a unimodal distribution with a mean pairwise sequence identity near ~25%.

C. Sector positions derived from the CMGC alignment (blue) or kinome-wide alignment (yellow) are distributed along the primary and secondary structure of the CMGC/MAPK ERK2. Subfamily-specific regions, such as the MAPK-insert, are only part of the sector derived from the CMGC alignment.

D. The relationship between the sector and positional conservation (computed as the Kullback-Leibler relative entropy, D_i) for both the CMGC and kinome-wide alignments. Sector positions

are highlighted in blue or yellow for the CMGC and kinome-wide alignments respectively. Red stars indicate highly conserved positions (defined as $D_i > 2.0$ in the kinome-wide alignment).

E. The kinome-wide and CMGC-specific sectors (yellow and blue transparent surfaces, respectively) mapped on human ERK2 (gray ribbon) (PDB: 2ERK). Conserved positions are shown as red spheres.

Supplementary Figure 3. ERK2 mutations within the kinase sector are enriched for loss-of-function.

A. The CMGC sector displayed as a transparent cyan surface overlaid by a ball-and-stick model of *R. norvegicus* ERK2 (PDB: 2ERK). Red-white-blue heat map color coding of the ball-and-stick model indicates residues that when mutated by Brenan et al. (29) led to inferred gain-of-function (GOF), neutral, and loss-of-function (LOF) activity, respectively, of human ERK2.

B. Fisher's exact table demonstrating statistically significant enrichment of inferred LOF mutations in ERK2 with CMGC sector-connected positions.

Supplementary Figure 4. The kinase sector encompasses the catalytic and regulatory spines.

A. The positions of the catalytic (C) and regulatory (R) spines as defined by Kornev et al (30, 31), yellow and dark red spheres respectively) are shown on a grey ribbon diagram of protein kinase A (PDB: 2CPK).

B. The kinome-wide sector is overlaid in blue on panel (A).

C. A vertical slice half way through panel (B) revealing the overlap between the spines and the sector. All spine positions are encapsulated within the sector or sector-connected.

Supplementary Figure 5. The relationship of negatively charged surface positions to the kinome-wide EPK sector. Within the main text, we show a statistically significant association

with the sector defined using the CMGC specific alignment. For comparison, here we show the results with the kinome-wide (EPK) alignment.

A. Space-filling model of Kss1 in gray, with the kinome-wide sector in blue. Positions with neutral, loss-of-function and gain-of-function mutations are color-coded (yellow, red, green, respectively).

B. Fisher's exact table demonstrating statistically significant enrichment of functional residues at sector-connected positions derived from the kinome-wide alignment.

METHODS

Yeast strains and plasmids

Yeast strains and plasmids used in this work are described in Supplementary Tables 1 and 2, respectively. All strains are in the W303 genetic background. Gene deletions were performed by one-step PCR as described (35). All Kss1 mutants were integrated into yeast genome as a single copy expressed from the endogenous *KSS1* promoter.

Site-directed mutagenesis

Site-directed mutagenesis was performed with QuickChange according to the manufacturer's directions (Agilent).

Cell growth and treatment with α factor

All cells were grown in synthetic complete media with dextrose (SDC). Three single colonies from each Kss1 strain bearing the *AGA1pr-YFP* reporter were inoculated in 1 ml SDC in 2 ml 96-well deep well plates and serially diluted 1:5 three times. Plates were incubated overnight at 30°C. In the morning cells from the row that had been diluted 1:25 were typically found to have OD₆₀₀ ~0.5. These cells were diluted 1:5 in 4 rows of a 96 well U-bottom micro-titer plate in a total volume of 180 μ l and incubated for 1 hour at 30°C. In each row, cells were treated with different concentrations of α factor: 0, 0.01, 0.1 and 1 μ M (10x stocks of α factor were prepared and 20 μ l were added to 180 μ l cells). Treated cells were incubated for an additional 4 hours at 30°C before translation was stopped by addition of 50 μ g/ml cycloheximide. Cells were incubated for an additional hour at 30°C to allow time for fluorophores to mature. For experiments with estradiol, everything is the same except that all media contained 20 nM estradiol for the duration of the overnight growth and throughout the experiment.

Flow cytometry

The AGA1pr-YFP reporter was measured by flow cytometry by sampling 10 μ l of each sample using a BD LSRFortessa equipped with a 96-well plate high-throughput sampler. Data were left ungated and FlowJo was used to calculate median YFP fluorescence. Bar graphs show the average of the median of the three independent colonies that were assayed, and error bars are the standard deviation.

Confocal microscopy

96 well glass bottom plates were coated with 100 μ g/ml concanavalin A in water for 1 hour, washed three times with water and dried at room temperature. 80 μ l of cells that had been treated with pheromone at the indicated concentrations for 3 hours were diluted to OD₆₀₀ ~0.05 and added to a coated well. Cells were allowed to settle and attach for 15 minutes, and unattached cells were removed and replaced with 80 μ l SDC media. Imaging was performed at the W.M Keck Microscopy Facility at the Whitehead Institute using a Nikon Ti microscope equipped with a 100 $\times$, 1.49 NA objective lens, an Andor Revolution spinning disc confocal setup and an Andor EMCCD camera. Images were analyzed in ImageJ.

Immunoprecipitation of 3xFLAG-tagged Kss1 and mutants

2 x 250 ml cultures of each strain were grown to OD₆₀₀=0.8 at 30°C with shaking, one in SDC and the other in SDC + 20 nM estradiol. The SDC culture was left untreated while the SDC + estradiol culture was treated with 1 μ M alpha factor for 30 minutes. Samples were collected by filtration and filters were snap frozen in liquid N₂ and stored at -80°C. Cells were lysed frozen on the filters in a coffee grinder with dry ice. After the dry ice was evaporated, lysate was resuspended in 1 ml IP buffer (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% triton x-100, 0.1% DOC, complete protease inhibitors), transferred to a 1.5 ml tube and spun to remove cell debris. Clarified lysate was transferred to a fresh tube and serial IP was performed. First, 25 μ l of anti-FLAG magnetic beads (50% slurry, Sigma) were added, and the mixture was

incubated for 2 hours at 4°C on a rotator. Beads were separated with a magnet and the supernatant was removed. Beads were washed 5 times with 1 ml IP buffer and bound material eluted 2x with 25 µl of 1 mg/ml 3xFLAG peptide (Sigma) in IP buffer by incubating at room temperature for 10 minutes. Beads were separated with a magnet and the two eluates were pooled in a fresh tube. 10 µl eluate was analyzed by Western blotting.

Western blotting

Total protein was TCA purified from cells as described (36). 10 µl of each sample was loaded into 4-15% gradient SDS-PAGE gels (Bio-Rad). Gels were run at 25 mA for 45 minutes, and blotted to PVDF membrane at 225 mA for 40 minutes. After 1hr blocking in Li-Cor blocking buffer, membranes were incubated with anti-FLAG primary antibody (SIGMA, F3165), anti-phospho-PKA substrate, anti-phospho p44/42 (Cell Signaling, 9101), and/or anti-PGK (22C5D8) overnight at 4°C on a platform rotator (all 1:1000 dilutions in blocking buffer). Membranes were washed three times with TBST and probed by anti-mouse or anti-rabbit IR dye-conjugated IgG (Li-Cor, 926-32352, 1:10000 dilution). The fluorescent signal was detected with the Li-Cor/Odyssey system.

Statistical Coupling Analysis (SCA)

SCA was performed as described in (24) using PySCA 6 (<http://reynoldsk.github.io/pySCA/>) for two different multiple sequence alignments of the kinase catalytic domain: one specific to the CMGC kinases (635 sequences), and one containing 7128 kinases sampled across the kinome. The CMGC alignment was constructed by searching kinbase (<http://kinase.com/kinbase/>). Sequences were filtered for a length of 250-350 amino acids, and aligned by Promals3D (37) including the PDBS: 2B9H, 1BI8, 1Q97, 2ERK, 2F49, 2F9G, 2IW8, 2R7I, as reference structures. The kinome-wide alignment was previously constructed by the Shokat lab and was downloaded from <http://sequoia.ucsf.edu/ksd/> (38). Following alignment processing and the

application of sequence weights (as described in (24)), the alignments contained 464 and 380 total effective sequences for the CMGC and EPK alignments respectively. For both alignments, we followed an identical procedure for defining the sector. Briefly, we compute a conservation-weighted covariance matrix between all pairs of amino acid positions (see Supplementary Text for discussion of the relationship between the sector, conservation, and allosteric hotpots). This matrix provides a statistical description of the "evolutionary coupling" between all pairs of amino acid positions. We then analyze this matrix by conducting principle components analysis (PCA), and rotating the top eigenmodes using independent components analysis (ICA). The top independent components are used to define sectors. For both kinase alignments, we define a single sector that includes all positions contributing to the top 4 independent components (ICs). The group of positions contributing to each IC groups is defined by fitting an empirical statistical distribution to the ICs and choosing positions above a defined cutoff (default, > 95% of the CDF). The full analysis of both families can be downloaded from github.

Defining sector-connected solvent accessible surface sites

We computed the relative solvent accessible surface area (RSA) over a homology model of Kss1 (34) using Michel Sanner's MSMS with a probe size of 1.4 Å, excluding all water and heteroatoms (39). A cutoff of 20% RSA was used to define solvent exposed surface positions (10). "Sector-connected" is defined as a position where any atom is within 4.0Å of a sector position.

SUPPLEMENTARY TEXT

The relationship between the sector, conservation, and allosteric hotpots

In this work we analyze the statistical association between the sector and functional measurements of mutational effects from three different datasets: (1) saturation mutagenesis of ERK2 (6,810 mutations across 359 positions, Supplementary Fig. 3 and Supplementary Table 5) (29), (2) an alanine scan of negatively charged positions on the Kss1 surface (40 mutations, Fig. 4a,b, Supplementary Fig. 5, Supplementary Table 6) and (3) functional mutations across a diversity of kinases surveyed from the literature (78 mutations mapped to 45 unique sites, Fig. 4c,d, Supplementary Table 7-8). The goal of this section is to provide a more complete discussion of the sector definition, as well as the relationship between the sector and these experimental datasets.

The sector is defined using the top four independent components (ICs) of the so-called “SCA matrix” ($\tilde{C}_{ij}$), a conservation-weighted covariance matrix between all pairs of amino acid positions (22, 24). In our analysis of the protein kinases, we group all of the positions contributing to the top ICs into a single sector, with the rationale that this is the most conservative interpretation in the absence of experimental data indicating functional or structural independence between residue groups. Though distinct from the goals of this paper, a further analysis of how parts of the sector may diverge in particular kinase subfamilies and how these residue groups relate to the evolutionary tuning of different biochemical properties is interesting, and addressed separately in concurrent work from Creixell et al (manuscript in prep.).

The conservation weighting of amino acid correlations is a defining feature of SCA and is applied with two complementary goals in mind: (1) to emphasize co-evolution between conserved (and thus likely functionally relevant) positions and (2) to minimize the contribution of purely phylogenetic correlations that are expected to emerge at weakly conserved positions. The origin of the conservation weights is described more completely in (24), but the weights are

applied as: $\tilde{C}_{ij}^{ab} = \frac{\partial D_i^a}{\partial f_i^a} \frac{\partial D_j^b}{\partial f_j^b} C_{ij}^{ab}$, where D_i is the Kullback-Leibler relative entropy, and C_{ij}^{ab} is the unweighted covariance between the frequencies (f) of a pair of amino acids a, b at a particular pair of positions i, j . Thus, we expect some redundancy between the information captured by sector positions (defined from the conservation weighted pairwise correlations) and simpler measures like the single-site conservation (D_i) (40). Accordingly, we consider the statistical association of conservation alone with the functional mutagenesis data alongside our analysis of sector positions.

We computed the statistical association between functional mutations and either: (1) the sector, at several cutoffs ($p = 0.95, 0.96, 0.97$ and 0.98) and (2) conserved positions, at several cutoffs ($D_i = 1, 1.15, 1.3$ and 2). This was done for two alignments: one containing only the CMGC family kinases, and one encompassing kinases across the full kinome (referred to as the EPK alignment). The cutoffs for the sector were chosen to span a range of sector sizes (e.g. from 41-91 amino acid positions, ~15-33% of the kinase). The cutoffs for conservation (D_i) were chosen to give similar numbers of amino acid positions as the sector cutoffs (Supplementary Table 5). These cutoffs include amino acid positions spanning “moderate” to more “stringent” levels of conservation: to map D_i to a more easily interpreted measure we computed the frequency of the most conserved amino acid at each position included in the cutoff. For the EPK alignment, we see that the D_i cutoffs of 1, 1.15, 1.3 and 2 correspond to conserved amino acid frequencies of 0.23, 0.39, 0.39 and 0.68 respectively. Following the definition of sector and conserved positions, we used a one-tailed Fisher exact test on a two-by-two contingency table (as in Fig. 4b) to evaluate the probability that the observed association between the sector and experimental data (or any association more extreme) is obtained randomly. We found that both the sector positions and conserved positions have a statistically significant association with the functional data over a range of cutoffs ($p < 0.05$, Supplementary Tables 5,6,8). This observation

forms the core of the argument that specific, evolutionarily conserved and co-evolving positions act as allosteric hotspots on the protein surface.

Though this statistical association does not depend strongly on the choice of alignment, we do observe some subtle differences. For example, the CMGC-specific alignment has a slightly better association with the ERK2 saturation mutagenesis data than the full alignment (Pearson χ^2 $p = 0.05$), while the full alignment performs better than the CMGC alignment when compared to the kinome-wide sampling of mutations (Pearson χ^2 $p = 4.7E-7$). In both cases, the sector definition agrees better with the experimental data when the underlying alignment is more representative of the kinases being compared.

The sector positions and conservation show a statistically equivalent association with the functional data (as assessed by comparing the two contingency tables by Pearson χ^2), meaning that it is difficult to distinguish between the functional significance of conserved residues and sector residues (40). However, the goal of this work is not to test the sector as an exclusive model for allosteric networks in proteins. Rather, our central claim is that allosteric potential is non-uniformly loaded into a handful of positions on the protein surface and that these facilitate the evolution of new regulation. The sector provides one way to identify these positions, and unlike single-site conservation, leads naturally to the interpretation that these positions form a cooperative network embedded within the protein structure.

Background on Kss1

The MAPK Kss1 is expressed in both haploid and diploid *S. cerevisiae* cells. Kss1 is activated via phosphorylation by Ste7 (MEK). When overproduced, Kss1 stimulates recovery from pheromone-imposed G1 arrest and was first identified as a suppressor of Sst2 mutations (16). Kss1 is also involved in filamentous (invasive) growth in haploid cells (17) and pseudohyphal development in diploid cells (20). While Kss1 is concentrated in the nucleus, stimulation with mating pheromone results in relocation of Kss1 to the cytoplasm (41, 42).

- 646 ▪ Length: 368 amino acids
- 647 ▪ Kinase domain: residues 13-313
- 648 ▪ ATP binding signature: residues 19-43
- 649 ▪ Kss1 forms an initial tight complex with the MEK Ste7 (K_D of ~5 nM) that is not the ES
- 650 conformation (43).
- 651 ▪ Residues on Kss1 phosphorylated by the MEK Ste7 within the activation loop: T183, Y185
- 652 ▪ The mutant K42R inactivates Kss1 activity, but does not affect phosphorylation of the
- 653 activation loop residues.
- 654 ▪ Kss1 binds to Ste7 (MEK), Ste12 (transcription factor), Dig1 (transcription regulator, Kss1
- 655 substrate), Dig2 (transcription regulator, Kss1 substrate) and other phosphorylation
- 656 substrates.
- 657 ▪ Kss1 exhibits both a kinase-dependent positive activity and a kinase-independent inhibitory
- 658 activity:
 - 659 ○ Kss1 positive activity requires both activation by Ste7 and Kss1 catalytic activity,
 - 660 ○ Kss1 inhibitory activity requires only the Kss1 protein

1 MARTITFDIP SQYKLVDLIG EGAYCTVCSA IHKPSGIKV A IKKIQPFSSK LFVTRTIRI
61 KLLRYFHEHE NIISILDKVR PVSIDKLNAY YLVEELMETD LQKVINNQNS GFSTLSDDHV
121 QYFTYQILRA LKSIHSAQVI HR IKPSNLL LNSNCDLKVC DFGCLARCLAS SSDSRETLVG
181 FMTEYVATRW YRAPHIMLTF QEYTTAMDIW SC CILAEMV SGKPLFPGRD YHHQLWLILE
241 VLGTSPFEDF NQIKSKRAKE YIANLPMRPP LPWETVWSKT DLNPDMDILL DKMLQFNPDK
301 RISAAEALRH PYLAMYHDPD DEPEYPLNL DDEFWKLDNK IMRPEEEEV PIEMLKDMLY
361 DELMKTME*

Kss1 protein sequence. 18 residues most highly conserved in protein kinases are highlighted in green or (K42 is highlighted blue).

Schematic representation of Kss1 kinase-dependent positive activities and kinase-independent inhibitory activities at pheromone-responsive gene promoters (A) and filamentation gene promoters (B).

Supplementary Table 1. Plasmids

pDP number	Nickname	Description	Marker
11	AGA1pr-YFP	pNH605-AGA1pr-YFP; single integration	<i>LEU2</i>
436	Kss1-3xFLAG-V5	pNH604-KSS1pr-KSS1-3xFLAG-V5; single integration; template for all Kss1 mutants	<i>TRP1</i>
1	GEM	pRS306-GAL4dbd-EstradiolReceptor-VP16ad; integrative	<i>URA3</i>
603	Ras2(G19V)	pNH603-GALpr-RAS2(G19V)	<i>HIS3</i>
333	YFP-Ras2(G19V)	pNH603-GALpr-YFP-RAS2(G19V)	<i>HIS3</i>

Supplementary Table 2. Yeast strains

DPY number	Nickname	Genotype
512	kss1Δ	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2
513	WT	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1-3xFLAG-V5
379	D8A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D8A)-3xFLAG-V5
380	D17A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D17A)-3xFLAG-V5
381	D21A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D21A)-3xFLAG-V5
382	E68A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E68A)-3xFLAG-V5
383	E70A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E70A)-3xFLAG-V5
384	D77A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D77A)-3xFLAG-V5
385	D85A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D85A)-3xFLAG-V5
386	E98A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E98A)-3xFLAG-V5
387	D117A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D117A)-3xFLAG-V5
388	D118A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D118A)-3xFLAG-V5
389	D156A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D156A)-3xFLAG-V5
390	D173A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D173A)-3xFLAG-V5
391	E176A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D176A)-3xFLAG-V5
392	E184A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D184A)-3xFLAG-V5
393	E202A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E202A)-3xFLAG-V5
394	D230A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D230A)-3xFLAG-V5
395	E248A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E248A)-3xFLAG-V5
396	E260A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E260A)-3xFLAG-V5
397	E274A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E274A)-3xFLAG-V5
398	D281A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D281A)-3xFLAG-V5
399	D285A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D285A)-3xFLAG-V5
400	D288A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D288A)-3xFLAG-V5
401	D291A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D291A)-3xFLAG-V5
402	D299A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D299A)-3xFLAG-V5
403	E306A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E306A)-3xFLAG-V5
404	D318A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D318A)-3xFLAG-V5
405	D321A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D321A)-3xFLAG-V5
406	E324A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E324A)-3xFLAG-V5
407	D331A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D331A)-3xFLAG-V5
408	D332A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D332A)-3xFLAG-V5
409	E333A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E333A)-3xFLAG-V5
455	D338A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D338A)-3xFLAG-V5
456	E345A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E345A)-3xFLAG-V5
457	E347A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E347A)-3xFLAG-V5
458	E348A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E348A)-3xFLAG-V5
459	E349A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E349A)-3xFLAG-V5
460	E353A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E353A)-3xFLAG-V5
461	D357A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(D357A)-3xFLAG-V5
462	E361A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E361A)-3xFLAG-V5
463	E362A	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 KSS1pr-KSS1(E362A)-3xFLAG-V5
674	pka-kss1Δ	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3
675	pka-Kss1 WT	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1-3xFLAG-V5::TRP1
677	pka-D8	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8D8)-3xFLAG-V5::TRP1
676	pka-A8	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8A8)-3xFLAG-V5::TRP1
678	pka-S8	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8S8)-3xFLAG-V5::TRP1
683	pka-E70	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8E70)-3xFLAG-V5::TRP1
682	pka-A70	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8A70)-3xFLAG-V5::TRP1
684	pka-S70	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8S70)-3xFLAG-V5::TRP1
680	pka-E68	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8E68)-3xFLAG-V5::TRP1
679	pka-A60	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8A60)-3xFLAG-V5::TRP1
681	pka-S68	W303A MATa ADE2 fus3Δ::KAN kss1Δ::HYG AGA1pr-YFP::LEU2 GEM::URA3 GALpr-RAS2(G19V)::HIS3 KSS1pr-KSS1(RRx8S68)-3xFLAG-V5::TRP1
321	Ras2(G19V)	W303A MATa ADE2 GEM::URA3 GALpr-RAS2(G19V)::HIS3
342	YFP-Ras2(G19V)	W303A MATa ADE2 GEM::URA3 GALpr-YFP-RAS2(G19V)::HIS3

Supplementary Table 3. Comparison of Kss1 point mutations from the literature with our data and the CMGC sector (cutoff p=0.95).

Mutation in this work	Sector connected	Mutation	Phenotype	Reference
		Kss1-del	Filamentous and invasive growth decreased	1
			Mates at a detectable frequency	
		Kss1 overexpression	Suppresses the recovery defect in Sst2 mutant cells	2
			Stimulates recovery from pheromone-induced G1 arrest	
yes		G20S	Filamentation-defective allele in kss1/kss1 null strain	3
yes		Y24F	In <i>pombe</i> Cdc2, phosphorylation and dephosphorylation of the tyrosine at this position is critical to control of the kinase	1
			Basal and pheromone-induced phosphorylation of Kss1 is not detectably altered	1
			Overproduction/Halo assay: worked like WT Kss1	1
			Represses invasive growth when Ste7 is absent	4
			Catalytically in active, but can be phosphorylated	4
			Represses invasive growth when Ste7 is present	
yes		G25R	Filamentation-defective allele in kss1/kss1 null strain	3
			Used to show Kss1-imposed repression at pheromone induced genes	5
yes		K42R	Inactivates kinase, leaves phosphorylation sites intact	6
			In the absence of fus3, no trx induction of fus1-lacZ	
			This first lysine aligns with an invariant lysine in all other protein kinases	1
			In every kinase examined mutation of this lysine abolishes detectable kinase activity	1
			Pattern of basal and pheromone-induced phosphorylation is indistinguishable from WT Kss1	1
			Mutation abolishes Kss1 function <i>in vivo</i>	1
			Unable to complement the mating deficiency of a kss1del, fus3del strain	1
			Overproduction/Halo assay: worked like WT Kss1	1
			Used to show that Ste7 is a substrate of Kss1 <i>in vitro</i> , K42R is the negative control	7
			Filamentation-defective allele in kss1/kss1 null strain	3
yes		K43R	Pattern of basal and pheromone-induced phosphorylation is indistinguishable from WT Kss1	1
			Overproduction/Halo assay: worked like WT Kss1	1
D117A	yes	D117E	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	N148S	Filamentation-defective allele in kss1/kss1 null strain	3
D156A	yes	D156E	Hyperfilamentation allele - class 1, proposed to either increase the kinase activity of Kss1 OR confer resistance to inactivating phosphatases	3
D156A	yes	D156G	Hyperfilamentation allele - class 1, proposed to either increase the kinase activity of Kss1 OR confer resistance to inactivating phosphatases	3
	yes	C160P	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	C160Y	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	D161N	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	V179A	Hyperfilamentation allele - class 2, residues proposed to be in the inhibitory function of Kss1	3
no		T183A	The lower band of the doublet is eliminated	1
			Required for Kss1 catalytic activity <i>in vivo</i>	1
			Overproduction/Halo assay: worked like WT Kss1	1
			Unactivatable	4
			Represses invasive growth when Ste7 is present	

	no	T183M	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	Y185F	In the absence of fus3, no trx induction of fus1-lacZ	6
			In the absence of fus3, no mating	1
			One of the phosphorylation sites on the activation loop	1
			Overproduction/Halo assay: worked like WT Kss1	1
			Phosphorylation is greatly reduced & the upper band of the doublet is eliminated	1
			Required for Kss1 catalytic activity in vivo	1
			Overproduction/Halo assay: worked like WT Kss1	1
			Filamentation-defective allele in kss1/kss1 null strain (manually constructed)	3
			Unactivatable	4
			Represses invasive growth when Ste7 is present	
	yes	V186A	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	Y203H	Filamentation-defective allele in kss1/kss1 null strain	3
	yes	I215V	Hyperfilamentation allele - class 1, proposed to either increase the kinase activity of Kss1 OR confer resistance to inactivating phosphatases	3
	yes	Y231C	Defective in binding to Ste12	4
	yes	D249G	Hyperfilamentation allele - class 2, residues proposed to be in the inhibitory function of Kss1	3
			Reduced co-immunoprecipitation of Ste12	3
			Lower enrichment of Kss1 in nucleus, weak relocation of Kss1 out of nucleus	8
E260A	no	E260G	Hyperfilamentation allele - class 2, residues proposed to be in the inhibitory function of Kss1	3
			Reduced co-immunoprecipitation of Ste12	3
D288A	yes	D288G	Hyperfilamentation allele - class 1, proposed to either increase the kinase activity of Kss1 OR confer resistance to inactivating phosphatases	3
D318A	no	D318A	Reduced ability to drive FUS1-lacZ, reduced Ste7 binding, aka kss1-7m1	9
			Decreased nuclear enrichment	8
D321A	no	D321A	Reduced ability to drive FUS1-lacZ, reduced Ste7 binding, aka kss1-7m2	9
	no	F334L	Filamentation-defective allele in kss1/kss1 null strain	3

Citations:

- 1) Ma D. et al. Phosphorylation and localization of Kss1, a MAP kinase of the *Saccharomyces cerevisiae* pheromone response pathway. *MCB* 6, 889-909 (1995)
- 2) Courchesne W.E. et al. A putative protein kinase overcomes pheromone-induced arrest of cell cycling in *S. cerevisiae*. *Cell* 58, 1107-1119 (1989)
- 3) Madhani H.D. et al. MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. *Cell* 91, 673-684 (1997)
- 4) Bardwell L. et al. Repression of yeast Ste12 transcription factor by direct binding of unphosphorylated Kss1 MAPK and its regulation by the Ste7 MEK. *Genes Dev* 12, 2887-2898 (1998)
- 5) Bardwell L. et al. Differential regulation of transcription: repression by unactivated mitogen-activated protein kinase Kss1 requires the Dig1 and Dig2 proteins. *PNAS* 95, 15400-15405 (1998)
- 6) Gartner A. et al. Signal transduction in *Saccharomyces cerevisiae* requires tyrosine and threonine phosphorylation of FUS3 and KSS1. *Genes Dev* 6, 1280-1292 (1992)
- 7) Bardwell L. et al. Signaling in the yeast pheromone response pathway: specific and high-affinity interaction of the mitogen-activated protein (MAP) kinases Kss1 and Fus3 with the upstream MAP kinase kinase Ste7. *MCB* 16, 3637-3650 (1996)
- 8) Pelet S. Nuclear relocation of Kss1 contributes to the specificity of the mating response. *Sci Rep* 7, 43636 (2017)
- 9) Kusari A.B. et al. A conserved protein interaction network involving the yeast MAP kinases Fus3 and Kss1. *J Cell Bio* 164, 267-277 (2004)

Supplementary Table 4. Sector positions mapped to several representative kinase structures

S. cerevisiae Pho85 (2PK9.pdb)

CMGC sector

7,14,16,17,18,19,21,22,24,33,34,35,36,37,47,49,52,53,57,62,64,66,68,71,79,82,83,84,88,89,111,116,120,124,125,131,132,133,134,135,136,138,140,145,148,151,152,153,154,155,156,166,168,169,171,172,173,174,175,176,177,178,180,186,191,193,196,197,198,200,201,208,209,210,211,217,221,226,227,228,233,234,270,272,282,285,291,294,295,296,297

EPK sector

7,14,16,17,18,19,21,24,33,34,35,36,50,52,53,57,62,64,66,68,70,72,79,80,81,82,83,85,87,88,89,118,120,121,123,124,125,129,131,132,133,135,136,138,139,140,148,151,152,153,154,155,171,173,174,177,178,189,191,193,194,195,196,200,201,277,278,282,285,286,294,295

S. cerevisiae Kss1 (KSS1.pdb - homology model)

CMGC sector

13,20,22,23,24,25,27,28,30,39,40,41,42,43,53,55,58,59,63,68,71,73,75,78,91,94,95,96,100,101,121,126,130,134,135,141,142,143,144,145,146,148,150,155,158,161,162,163,164,165,166,178,185,186,188,189,190,191,192,193,194,195,197,203,208,210,213,214,215,217,218,225,226,227,228,234,238,243,244,245,250,251,286,288,298,301,307,310,311,312,313

EPK sector

13,20,22,23,24,25,27,30,39,40,41,42,56,58,59,63,69,71,73,75,78,80,91,92,93,94,95,97,99,100,115,128,130,131,133,134,135,139,141,142,143,145,146,148,149,150,158,161,162,163,164,165,185,187,188,194,195,203,208,210,211,213,214,218,219,284,287,288

R. norvegicus ERK2 (2ERK.pdb)

CMGC sector

23,30,32,33,34,35,37,38,40,49,50,51,52,53,63,65,68,69,73,78,80,82,84,87,100,103,104,105,109,110,125,130,134,138,139,145,146,147,148,149,150,152,154,159,162,165,166,167,168,169,170,182,185,186,188,189,190,191,192,193,194,195,197,203,208,210,213,214,215,217,218,225,226,227,228,234,238,243,244,245,250,251,284,286,296,299,305,308,309,310,311

EPK sector

23,30,32,33,34,35,37,40,49,50,51,52,66,68,69,73,78,80,82,84,93,100,101,102,103,104,106,108,109,110,128,130,135,137,138,139,143,145,146,147,149,150,152,153,154,162,165,166,167,168,169,185,187,188,194,195,203,208,210,211,213,214,218,219,291,292,296,299,300,308,309

M. musculus PKA (2CPK.pdb)

EPK sector

43,50,52,53,54,55,57,60,69,70,71,72,88,90,91,95,100,102,104,106,109,111,117,118,119,120,121,123,125,126,128,147,149,154,156,157,158,162,164,165,166,168,169,171,172,173,181,184,185,186,187,188,198,200,201,207,208,215,220,222,223,225,226,230,231,272,273,277,280,281,294,295

Supplementary Table 5. Statistical association between the sector, conservation and ERK2 mutational data

Sector positions

sector cutoff:	0.95	0.96	0.97	0.98
CMGC alignment	2.3 E-19 (Npos = 91)	4.4 E-20 (Npos = 81)	1.4 E-20 (Npos = 62)	8.4 E-10 (Npos = 41)
EPK alignment	5.3 E-11 (Npos = 72)	5.7 E-11 (Npos = 61)	9.0 E-11 (Npos = 47)	2.8 E-4 (Npos = 19)

Conserved positions

Di cutoff:	1	1.15	1.3	2
CMGC alignment	2.8 E-16 (Npos = 116)	3.9 E-17 (Npos = 88)	2.4 E-18 (Npos = 76)	3.4 E-18 (Npos = 32)
EPK alignment	4.7 E-8 (Npos = 72)	4.2 E-8 (Npos = 60)	3.7 E-10 (Npos = 44)	1.5 E-10 (Npos = 23)

Supplementary Table 6. Statistical association between the sector, conservation and KSS1 D/E Surface Mutations

(italics indicates insignificant relationships)

Sector positions

sector cutoff:	0.95	0.96	0.97	0.98
CMGC alignment	9.83E-03	6.14E-03	<i>7.20E-02</i>	<i>1.38E-01</i>
EPK alignment	3.29E-02	2.07E-02	<i>1.90E-01</i>	<i>8.97E-01</i>

Conserved positions

Di cutoff:	1	1.15	1.3	2
CMGC alignment	1.52E-02	2.01E-02	2.01E-02	<i>5.04E-01</i>
EPK alignment	<i>1.03E-01</i>	3.28E-03	3.28E-03	8.50E-03

Supplementary Table 7. Functional Kinase Mutations (sampled across the kinome)

	Kinase	Regulatory Position or Mutation	Pos. in Pho85	Pos. in Kss1	Phenotype	Reference
CMGC	Fus3	T180	V169	T183	activation loop phosphorylation site	1
		Y182	T171	Y185	activation loop phosphorylation site	1
		I161L	A157	C167	"unlocking mutation" - make Fus3 more Kss1-like	2
	Pho85	Y18F	Y18	Y24	reduced kinase activity, abolishes cyclin interaction	1
		K36R	K36	K42	loss of kinase activity	1
		E53A	E53	E59	loss of kinase activity, abolishes cyclin interaction	1
	Cdk2	K9F	K12	L18	reduced phosphorylation by CAK	1
		T14A	T17	A23	two fold increase in activity	1
		Y15F	Y18	Y24	two fold increase in activity	1
		T39	--	F47	phosphorylation/regulation site	3
		K88E	K90	S116	reduced phosphorylation by CAK (w/ K89V)	1
	Cdk6	K89V	K91	D117	reduced phosphorylation by CAK (w/ K89E)	1
		T160A	E168	M182	loss of kinase activity	1
		L166R	Y174	T188	reduced phosphorylation by CAK	1
		Y24	Y18	Y24	phosphorylation site	1
		T177	V169	T183	phosphorylation site	1
		R31	L25	--	interface mutation (w/ p19-INK4) in familial melanoma	4
TK	Fes	K590E	K36	K42	loss of kinase activity	1
		R609E	S55	K61	important to Fes/SH2 interaction	5
		M704V	L154	L164	reduced autophosphorylation and activity	1
		Y713F	N163	--	reduces kinase activity by 90%	1
		V743M	I192	I209	reduced kinase activity	1
	Hck	S759F	Q217	K259	reduced autophosphorylation and activity	1
		K290E	K36	K42	loss of kinase activity	1
		E305A	E53	E59	loss of kinase activity	1
		D381E	D133	D143	loss of kinase activity	1
		Y411A	E168	M182	reduced catalytic activity	1
		Y522F	--	L308	constitutively activated kinase	1
Eukaryotic-like	PKNE	K45M	K36	K42	loss of kinase activity	1
		T50	D41	P46	phosphorylation site	1
		T59	S48	T54	phosphorylation site	1
		T170	--	--	phosphorylation site	1
		T175	T171	Y185	phosphorylation site	1
	PKND	T178	Y174	T188	phosphorylation site	1
		H79A	Y69	L76	dimer interface, influences activity	6
		Y81A	V71	V79	dimer interface, influences activity	6
		D138N	D133	D143	catalytically inactive mutation	6
Other/eIF2a	PKR	D289A	T29	P34	inhibits dimerization (and kinase activation)	7
		L315F	L60	--	activating in absence of dimer	7
		Y323A	I72	R80	inhibits dimerization (and kinase activation)	7
		Y404H	F123	S133	activating in absence of dimer	7
		K429R	K148	K158	activating in absence of dimer	7
		T446	V170	E184	phosphorylation site	7
		T487A	--	R229	inhibits interaction w/ eIF2a	7
TKL	Tak1	F495I	--	--	inhibits interaction w/ eIF2a	7
		K63W	K36	K42	loss of kinase activity	1
		T178	L154	L164	phosphorylation site	8
		T184	I160	S170	phosphorylation site	8
	B-RAF	S198	P177	P194	phosphorylation site	1
		V599E	--	--	gain of function, associated w/ melanoma	9
		R461I	K12	L18	gain of function, associated w/ melanoma	9
		G465A	G16	G22	kinase dead, associated w/ melanoma	9
		G468A	A19	G25	kinase dead, associated w/ melanoma	9
		G468E	A19	G25	kinase dead, associated w/ melanoma	9

		N580S	N138	N148	kinase dead, associated w/ melanoma	9
STE	Mek1	<i>P124S</i>	<i>E63</i>	<i>E70</i>	gain of function, associated w/ melanoma	10
		E203K	Q146	D156	gain of function, associated w/ melanoma	10
AGC	Grk2	G475I	--	--	effects receptor phosphorylation	11
		V477D	--	--	effects receptor phosphorylation	11
		I485D	--	--	effects receptor phosphorylation	11
	PKC α	F435C	Y112	Y122	kinase dead; endometrial cancer	12
		A444V	--	--	reduced kinase activity; endometrial cancer	12
		D481E	D151	D161	reduced kinase activity; colorectal cancer	12
	PKC β	G585S	S287	L289	reduced kinase activity; lung cancer	12
		Y417H	T79	Y91	kinase dead; liver cancer	12
		A509V	A176	A193	loss of physiological response; breast cancer	12
	PKC γ	A509T	A176	A193	kinase dead; colorectal cancer	12
		F362L	Y18	Y24	kinase dead; endometrial cancer	12
		G450C	V110	V120	kinase dead; endometrial cancer	12
	PKC δ	P524R	P177	P194	loss of physiological response; pancreatic cancer	12
		D530G	--	M207	loss of physiological response; colorectal cancer	12
		P568A	--	N264	kinase dead; head and neck cancer	12
	PKC η	K591E	L280	--	reduced kinase activity; breast cancer	12
		R596H	R285	I287	loss of physiological response; colorectal cancer	12
		G598V	S287	L289	loss of physiological response; lung cancer	12
	PKC ζ	E421K	D178	E195	loss of physiological response; breast cancer	12
CAMK	CHK2	S428F	--	--	unknown, associated with breast cancer	13
	CamKII	T286D	--	--	gain of function, associated with breast cancer	14

78 total mutations 45 unique sites

Citations:

- 1) Boutet E. et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. *Methods Mol. Biol* 1374, 23-54 (2016)
- 2) Good M. et al. The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. *Cell* 136, 1085-97 (2009)
- 3) Maddika S. et al. Akt-mediated phosphorylation of CDK2 regulates its dual role in cell cycle progression and apoptosis. *J Cell Sci* 121, 979-988 (2008)
- 4) Russo A. et al. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumor suppressor p16INK4a. *Nature* 395, 237-243 (1998)
- 5) Filippakopoulos P. et al. Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. *Cell* 134, 793-803 (2008)
- 6) Greenstein A. et al. Allosteric activation by dimerization of the PknD receptor Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *JBC* 282, 11427-11435 (2007)
- 7) Dey M. et al. Mechanistic link between PKR dimerization, autophosphorylation and substrate recognition. *Cell* 122, 901-913 (2005)
- 8) Yu Y. et al. Phosphorylation of Thr-178 and Thr-184 in the TAK1 T-loop is required for interleukin (IL)-1-mediated optimal NF κ B and AP-1 activation as well as IL-6 gene expression. *JBC* 283, 24497-24505 (2008)
- 9) Wan P.T. et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* 116, 855-867 (2004)
- 10) Nikolaev, S.I. et al. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nature Genetics* 44, 133-139 (2012)
- 11) Beaudrait A. et al. Mapping the putative GPCR docking site on GPCR kinase 2. *JBC* 289, 25262-25275 (2014)
- 12) Antal C.E. et al. Cancer-associated protein kinase C mutations reveal kinase's role as tumor suppressor. *Cell* 160, 489-502 (2015)
- 13) Shaag A. et al. Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. *Hum Mol Genet* 14, 555-563 (2005)
- 14) Chi et al. Phosphorylation of calcium/calmodulin-stimulated protein kinase II at T286 enhances invasion and migration of human breast cancer cells. *Scientific Reports* 6, 33132 (2016)

Supplementary Table 8. Statistical association between the sector, conservation and functional mutations sampled across a diversity of kinases.

Sector positions

sector cutoff:	0.95	0.96	0.97	0.98
CMGC alignment	2.50E-02	8.65E-03	3.06E-02	1.05E-02
EPK alignment	6.82E-04	3.41E-04	7.67E-04	2.79E-02

Conserved positions

Di cutoff:	1	1.15	1.3	2
CMGC alignment	5.09E-03	7.79E-03	4.41E-02	4.08E-02
EPK alignment	2.69E-04	4.19E-04	1.04E-03	2.48E-04

Figure 1
Pincus et al.

Figure 2
Pincus et al.

Figure 3
Pincus et al.

Figure 4
Pincus et al.

Figure S1
Pincus et al.

Figure S2
Pincus et al.

Figure S3
Pincus et al.

Figure S4
Pincus et al.

Figure S5
Pincus et al.