**Supplementary Tables**

**Table S1. Classification of mutations and definition of different classes.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mutations  | Classification parameter | Class | Class name | # of mutations |
| 103 somatic cancer mutations from COSMIC database V68 “All Cancer” | Frequency of observed samples | One sample | Single | 57 |
| Two or more samples | Reccurent | 46 |
| Zygosity | Homozygous | Homo | 27 |
| Heterozygous | Hetero | 49 |
| Types of cancer | Leukemia | Leukemia | 50 |
| Sarcoma | Sarcoma | 28 |
| The involvement of mutations in Zn-Coordination | Zn-Coordination | Zn | 24 |
| Not involved in Zn-Coordination | NonZn | 79 |
| All possible single nucleotide substitutions |  | Reference set | Random | 2102 |

**Table S2. P-values for different null hypotheses calculated by two-sample *t*-test.**

**P-values of less than 0.01 indicates that mean values of distributions of changes in unfolding (binding) energies for different classes of mutations are statistically significantly greater than random mutations.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P-values | Single | Recurrent | Homo | Hetero | Leukemia | Sarcoma | Zn | NonZn | All Cancer |
| **nCBL** | 0.034 | 0.000 | 0.000 | 0.024 | 0.000 | 0.314 | 0.000 | 0.233 | 0.000 |
| **CBL-S** | 0.019 | 0.001 | 0.001 | 0.050 | 0.000 | 0.478 | 0.000 | 0.368 | 0.000 |
| CBL-E2-S | 0.452 | 0.000 | 0.004 | 0.104 | 0.000 | 0.502 | 0.000 | 0.209 | 0.001 |
| pCBL-E2-S | 0.014 | 0.000 | 0.002 | 0.013 | 0.000 | 0.265 | 0.000 | 0.001 | 0.000 |

**P-values of less than 0.01 indicate that mean values of distributions of changes in unfolding (binding) energies for “Recurrent”, “Homo”, “Leukemia” and “Zn” mutations are statistically significantly greater than “Single”, “Hetero”, “Sarcoma” and “NonZn” mutations, respectively.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P-values | Recurrent/Single | Homo/Hetero | Leukemia/Sarcoma | Zn/NonZn |
| **nCBL** | 0.003 | 0.024 | 0.000 | 0.000 |
| **CBL-S** | 0.024 | 0.023 | 0.000 | 0.000 |
| CBL-E2-S | 0.000 | 0.021 | 0.003 | 0.000 |
| pCBL-E2-S | 0.002 | 0.035 | 0.001 | 0.012 |

**Table S3. Top 25% cancer mutations with the largest energy changes for each state in CBL activation cycle.**

|  |  |
| --- | --- |
| *ΔΔG*fold | *ΔΔG*bind |
| *nCBL* | ***CBL-S*** | ***CBL-E2-S*** | ***pCBL-E2-S*** |
| C381Y #&  | C416W #& | **C404R** #+ ! | W408S #& ! |
| C381R #+  | **C381Y**  #& | **C384R** #& ! | **C404Y**  #+ ! |
| C416Y #+ | C416Y #+ | G413R # ! | W408R #+ ! |
| C416W #& | **C381R**  #+ | **C381R** #+ ! | **C384R** #& ! |
| C404Y #+ | S376F \*+ | **C404Y** #+ ! | **C404R** #+ ! |
| C416R \* | **C416R** \* | W408R #+ ! | W408C #+ ! |
| C396Y #+ | **C404Y**  #+ | W408S #& ! | G413R # ! |
| G415V \*+ | C396Y #+ | **C381G** \*&  | **C381R** #+ |
| C401R #& | **L405P** \*& | **C401R** #& | W408L \* ! |
| H398Y \*+ | **M400R** \* | **C381Y** #& | G375P \* ! |
| C401Y # | C401Y # | **C384Y** #& ! | R420G \* ! |
| C401F \* | H398P \*+ | C404S #+ ! | R420P #+ ! |
| M400R \* | **C401R** #& | **C416R** \* | **C381G**  \*& |
| C404R #+ | **C401F** \* | C401S # | **C401R** #& |
| C381G \*& | **C404R** #+ | W408C #+ ! | **C416R** \* |
| C384Y #& | **C381G** \*& | **L405P** \*& | G375S \* ! |
| H398P \*+ | C396R #+ | **C401F** \* | **C384Y**  #& ! |
| G415S #& | **C384Y**  #& | **M400R** \* | C404S #\* ! |
| C384R #& | H398Y \*+ | C401Y # | R420L #& ! |
| H398Q \*& | C416S \* | Y371S \*+ | P417R #\* ! |
| C416S \* | C396G \*+ | F418S #+ ! | F418S #\* ! |
| H398R #& | H398Q \*& | Y371D #+ | **M400R** \* |
| C419R #& | **C384R** #& | I98T \*& | **C381Y** #& |
| L405P \*& | H398R #& | P417R #+ ! | P417S \* ! |
| G413R # | C419R #& | C416Y #+ | **L405P** \*& |
| Y371D #+ | Y371D \*+ | Y371N # | **C401F** \* |

“\*”: “Single”; “#”: “Recurrent”; “+”: “Homo”; “&”: “Hetero”; “!”: Interface; Bold: Mutations appear in all states. Each mutation is listed from the largest to smallest energy changes.

**Table S4. Binding free energy for wild-type CBL-E2-S and pCBL-E2-S states.**

|  |  |
| --- | --- |
|  | Δ*Gbind*(kcal mol-1) |
| **CBL-E2-S** | -14.20 |
| **pCBL-E2-S**  | -15.13 |

**Table S5. Pearson correlation coefficients (R) between** **experimental densitometryvalues and predicted changes in stability and binding affinity.** For comparison, several alternative methods are listed which predict the effects of mutations on protein function (PROVEAN, PolyPhen-2, MutationAssessor and InCa) and on unfolding free energy (Eris, Rosetta and PoPMuSiC) for 15 experimentally tested mutants.The change in absorption between wild-type and mutant proteins is measured as D = Dm/Dw (D – densitometry) which refers to a fraction of active mutants. The relationship between densitometry and changes in energy can be described by the Boltzmann equation relating the probability of a state with the energy of this state as D ~ exp (- (ΔGmut –ΔGWT)/RT). P-values were calculated to test the hypothesis about the equality of correlation coefficients to zero. If P-value is less than 0.01 or 0.05 the correlation is statistically significant. For alternative methods we list their best linear correlation coefficients.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *ΔΔGfold (kcal mol-1)**Boltzmann model* | *ΔΔGfold (kcal mol-1)**Linear model* | *ΔΔGbind (kcal mol-1)**Linear model* | *PROVEAN* | *PolyPhen-2* | *MutationAssessor*  | *InCa* | *Eris* | *Rosetta* | *PoPMuSiC* |
| nCBL | CBL-S | Max | nCBL | CBL-S | Max | CBL-E2-S | pCBL-E2-S | Max | nCBL | nCBL | nCBL |
| R | 0.78 | 0.77 | 0.83 | 0.69 | 0.69 | 0.76 | 0.31 | 0.48 | 0.45 | 0.58 | 0.04 | 0.40 | 0.12 | 0.50 | 0.21 | 0.52 |
| P-value  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.07 | 0.09 | 0.02 | 0.88 | 0.14 | 0.68 | 0.06 | 0.46 | 0.05 |

**Table S6. True positive rates are listed corresponding to different false positive rates.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ΔΔ*Gfold (kcal mol-1)* | ΔΔ*Gbind (kcal mol-1)* | *PROVEAN* | *PolyPhen-2* | *MutationAssessor* | *InCa* |
|  | nCBL | CBL-S | CBL-E2-S | pCBL-E2-S |
| 5% | 0.22 | 0.23 | 0.19 | 0.25 | 0.25 | NS | 0.33 | 0.46 |
| 10% | 0.31 | 0.28 | 0.28 | 0.35 | 0.31 | NS | 0.34 | 0.52 |
| 20% | 0.46 | 0.37 | 0.33 | 0.45 | 0.39 | NS | 0.36 | 0.60 |

103 cancer mutations are assumed to be positive; 2102 random mutations are assumed to be negative. True positive rate (TP/(TP + FN)) is defined as the number of true positive (TP) divided by the sum of true positive and false negative (FN). False positive rate (FP/(FP+TN)) is defined as the number of false positive (FP) divided by the sum of false positive and true negative (TN).

**Table S7. Experimental relative densitometry values for HEK293T cell line, predicted changes in stability (ΔΔ*G*fold kcal mol-1), binding affinity (ΔΔ*G*bind kcal mol-1) and comparison with other methods for 15 experimentally tested mutations.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutations | Densitometry | Stability | Binding affinity | PROVEAN | PolyPhen-2 | MutationAssessor | InCa |
| nCBL | CBL-S | CBL-E2-S | pCBL-E2-S |
| C396R | **0.03±0.007** | **4.65** | **8.57** | **0.98** | **0.99** | **High** | **High** | **High** | **High** |
| H398Q | **0.04±0.019** | **7.28** | **7.34** | 0.79 | **1.00** | **High** | **High** | **High** | **High** |
| Y371H | **0.06±0.038** | **3.58** | **3.43** | **0.98** | **1.04** | **High** | **High** | *Med* | **High** |
| K382E+ | **0.07±0.016** | **2.30** | 0.56 | 0.82 | **1.11** | **High** | **High** | *Med* | **High** |
| C381A | **0.09±0.028** | **8.36** | **7.55** | **2.29** | **1.83** | **High** | **High** | **High** | *Med* |
| *L399V* | *0.24±0.051* | 1.64 | 0.22 | 0.80 | 0.90 | **High** | **High** | *Med* | Neu |
| *G375P+* | *0.27±0.094* | -0.19 | **5.00** | 0.67 | **2.09** | **High** | **High** | *Med* | **High** |
| *P395A* | *0.46±0.138* | **1.97** | **2.93** | 0.66 | **1.15** | **High** | **High** | *Med* | Neu |
| *V391I* | *0.56±0.185* | -0.14 | -0.01 | 0.81 | 0.79 | Neu | Neu | Neu | **High** |
| M374V+ | 0.88±0.104 | 1.36 | 0.87 | 0.81 | 0.90 | **High** | **High** | *Med* | Neu |
| V430M | 1.03±0.150 | -0.04 | -1.15 | 0.87 | **1.00** | Neu | **High** | *Med* | Neu |
| P428L | 1.04±0.059 | 0.80 | 0.99 | 0.85 | 0.70 | **High** | **High** | *Med* | *Med* |
| S80N | 1.08±0.115 | -0.42 | -0.5 | 0.71 | 0.86 | **High** | **High** | *Med* | **High** |
| H94Y | 1.08±0.115 | -1.00 | -0.2 | 0.80 | 0.77 | **High** | **High** | *Med* | **High** |
| Q249E | 1.33±0.077 | 0.88 | 1.16 | 0.79 | 0.86 | **High** | **High** | *Med* | **High** |
| Cutoff |  | 1.80 | 2.04 | 0.87 | 0.95 |

Cutoff: is derived from the distribution of random mutations and is equal to the mean value plus standard error. Mutation names are in bold, italic or regular font if mutants abolished, attenuated or did not affect ligase activity respectively. “+” indicates mutant sites located on CBL-E2 interface of active state. Experimental data for S80N and H94Y mutations are derived from double mutation of S80N/H94Y.