Supplementary figures:

Figure S1:

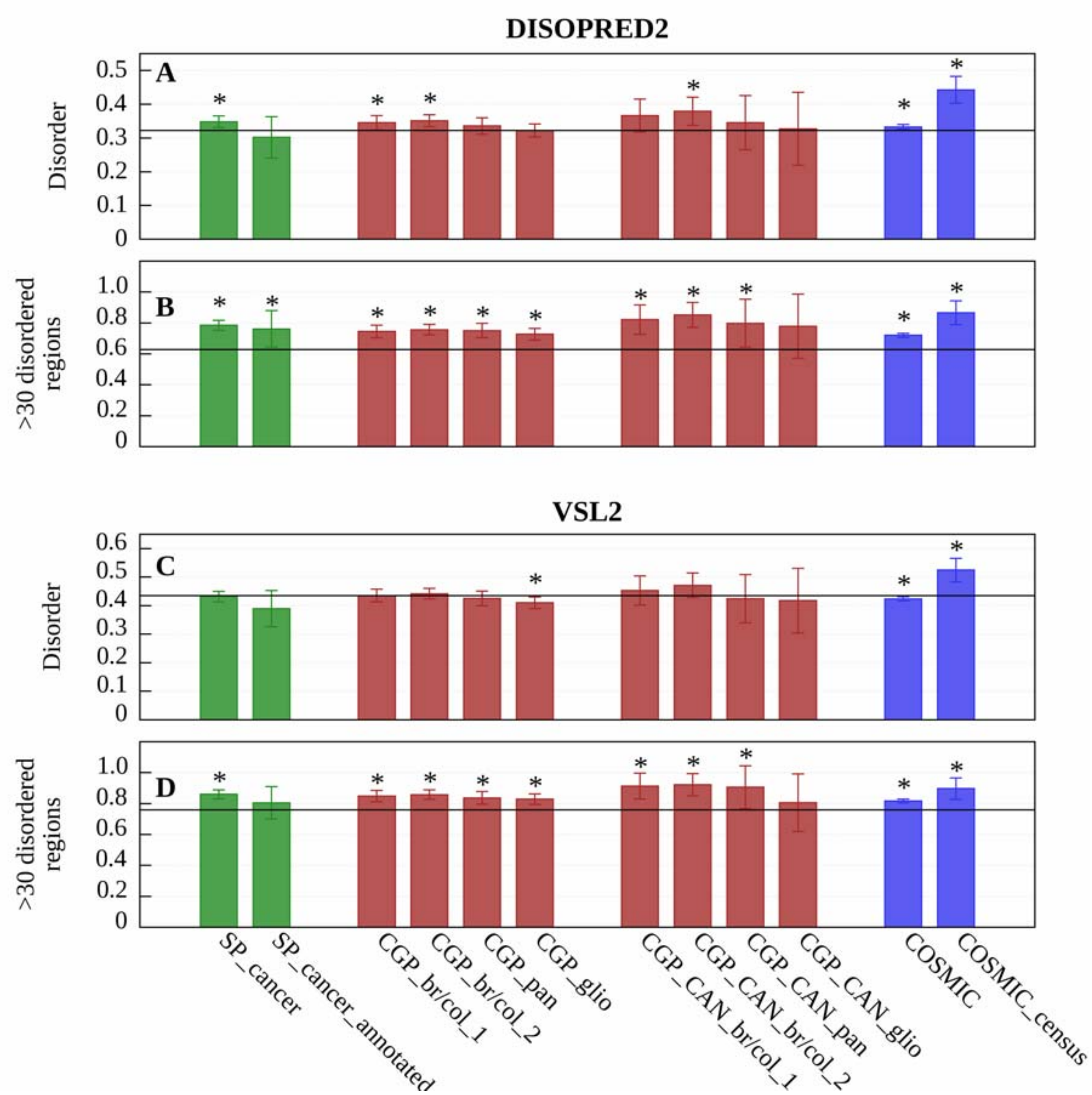

Figure S1: Average ratio of disordered residues (a and c) and ratio of proteins containing >30 residue long disordered regions (b and d) in the 12 datasets analyzed calculated with DISOPRED2 and VSL2. Black horizontal lines represent the average values obtained for the proteins of the human proteome taken from SwissProt. Flags show the confidence interval of α=0.01 calculated from the standard error of the mean of random selected samples from the human proteome (see Data and methods). Significant differences are marked with an asterisk.

Figure S2:

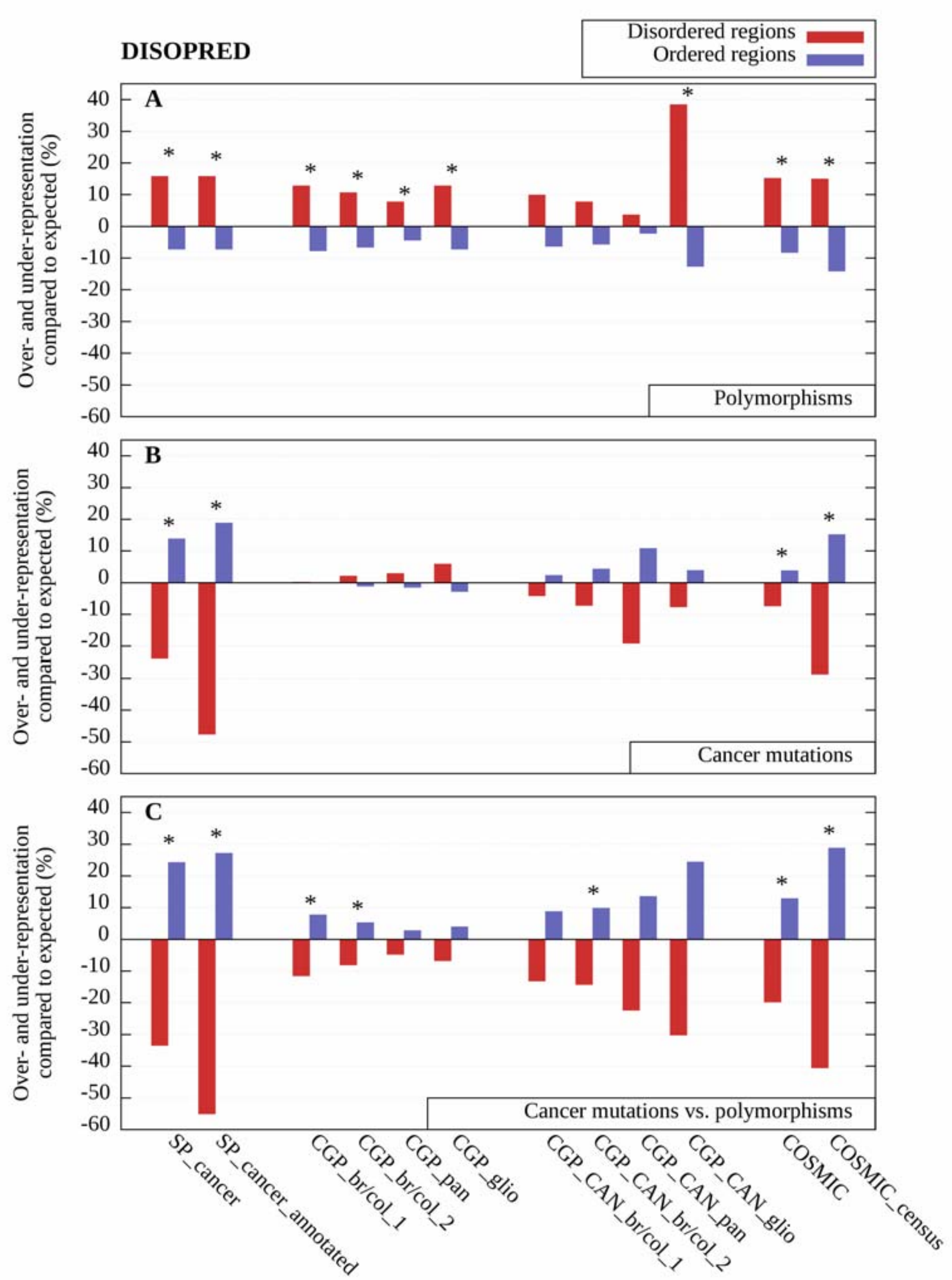

Figure S2: Over- and under-representation of mutations in disordered (red) and ordered regions (blue) as compared to background distributions calculated with DISOPRED2 (see Data and Methods). (a) the distribution of polymorphisms as compared to the uniform random distribution; (b) the distribution of cancer-associated mutations as compared to the uniform random distribution and (c) the distribution of cancer-associated mutations as compared to the expected values based on the distribution of polymorphisms shown in (a). Significant differences are marked with an asterisk.

Figure S3:

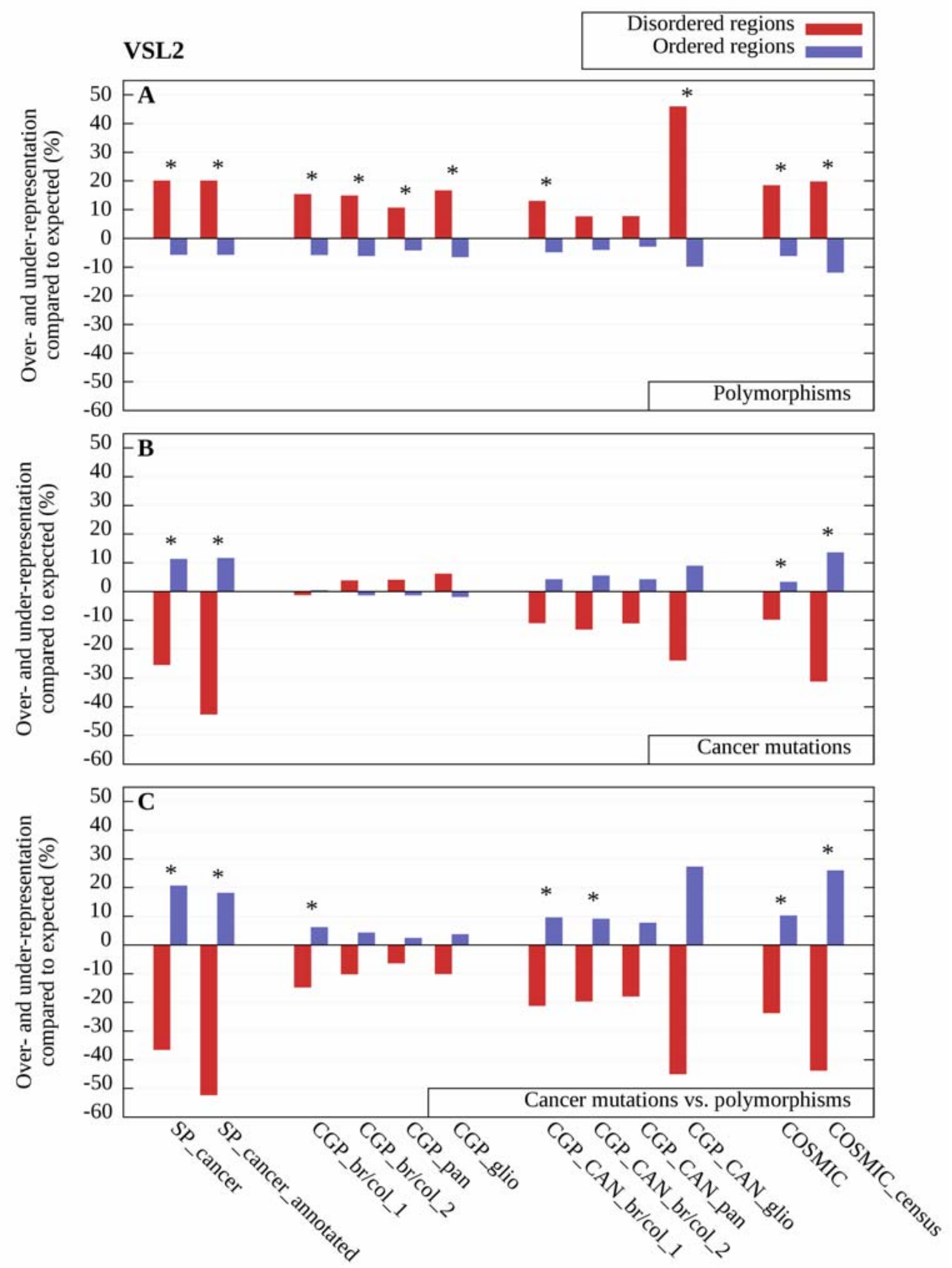

Figure S3: Over- and under-representation of mutations in disordered (red) and ordered regions (blue) as compared to background distributions calculated with VSL2 (see Data and Methods). (a) the distribution of polymorphisms as compared to the uniform random distribution; (b) the distribution of cancer-associated mutations as compared to the uniform random distribution and (c) the distribution of cancer-associated mutations as compared to the expected values based on the distribution of polymorphisms shown in (a). Significant differences are marked with an asterisk.

Supplementary tables:

Table S1: Average ratio of disordered residues, ratio of proteins containing >30 residue long disordered regions and length in the 12 datasets analyzed and in the Human proteome used as a reference. Disorder values were calculated with IUPred. The table also shows the standard error of the mean of the reference values and the probability of the respective database averages coming from the background distribution (see Data and methods).

Table S2: Observed and expected number of cancer-associated mutations and polymorphisms in disordered and ordered regions in the 12 datasets analyzed. Tables show the expected values of mutations calculated from the randomly uniform distribution ("Cancer mutations") and calculated from the distribution of polymorphisms ("Cancer mutations compared to polymorphisms"). The quoted p-values show the significance of the difference between observed and expected values and were obtained using the χ^2 test.

Table S3: Observed and expected number of cancer-associated mutations and polymorphisms in disordered binding regions, disordered and ordered regions in the 12 datasets analyzed. Tables show the expected values of mutations calculated from the randomly uniform distribution ("Cancer mutations") and calculated from the distribution of polymorphisms ("Cancer mutations compared to polymorphisms"). The quoted p-values show the significance of the difference between observed and expected values and were obtained using the chi^2 test.

Table S4: Bin boundaries for 5 features calculated on the COSMIC census database and the human proteome.

Table S1

Database name	Number of proteins	Ratio of disordered residues				Ratio of proteins with 30+ disordered segments				Length			
		average in database	average in human proteome	standard error	p-value	average in database	average in human proteome	standard error	p-value	average in database	average in human proteome	standard error	p-value
SP_cancer	1403	0.2423	0.2278	0.0065	0.02575	0.5802	0.4467	0.0133	$9.187*10^{-24}$	891.5	557.7	16.1	$1.509*10^{-95}$
SP_cancer_annot	113	0.1918	0.2278	0.0233	0.1211	0.5044	0.4467	0.047	0.2224	811.4	557.7	55.4	$4.432*10^{-6}$
CGP_br/col_1	924	0.2438	0.2278	0.0082	0.05238	0.5703	0.4467	0.0163	$3.732*10^{-14}$	861	557.7	20	$6.507*10^{-52}$
CGP_br/col_2	1335	0.2504	0.2278	0.0068	$9.330*10^{-4}$	0.594	0.4467	0.0134	$4.769*10^{-28}$	998.1	557.7	16.5	$8.033*10^{-157}$
CGP_pan	711	0.24	0.2278	0.0093	0.1936	0.5738	0.4467	0.0186	$9.104*10^{-12}$	1082.5	557.7	22.4	$2.114*10^{-121}$
CGP_glio	1089	0.2233	0.2278	0.0075	0.5485	0.5427	0.4467	0.015	$1.554*10^{-10}$	986.4	557.7	18.2	$1.255*10^{-122}$
CGP_CAN_br/col_1	174	0.2726	0.2278	0.0191	0.01827	0.6494	0.4467	0.0379	$8.322*10^{-8}$	1170.9	557.7	45.7	$4.617*10^{-41}$
CGP_CAN_br/col_2	243	0.2898	0.2278	0.0159	$1.002*10^{-4}$	0.6872	0.4467	0.032	$5.913*10^{-14}$	1226.8	557.7	39.1	$1.250*10^{-65}$
CGP_CAN_pan	64	0.2467	0.2278	0.0308	0.5353	0.625	0.4467	0.0625	$4.236*10^{-3}$	1130	557.7	73.8	$8.493*10^{-15}$
CGP_CAN_glio	36	0.2419	0.2278	0.0416	0.7339	0.6111	0.4467	0.0837	0.04884	1195.3	557.7	101.5	$3.175*10^{-10}$
COSMIC	8957	0.2314	0.2278	0.0026	0.1676	0.5195	0.4467	0.0053	$5.848*10^{-43}$	770.2	557.7	6.4	$1.077*10^{-241}$
COSMIC_census	261	0.3414	0.2278	0.0153	$1.085*10^{-13}$	0.6897	0.4467	0.0305	$1.463*10^{-15}$	912.4	557.7	37.3	$1.907*10^{-21}$

Table S2

SP_cancer

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	8,071	9,696	20.1%
Ordered	28,512	26,887	-5.7%

$p = 2.713*10^{-93}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	1,615	1,205	-25.4%
Ordered	3,631	4,041	11.3%

$p = 1.430*10^{-34}$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	1,897	1,205	-36.5%
Ordered	3,349	4,041	20.7%

$p = 4.422*10^{-88}$

SP_cancer_annotated

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	8,071	9,696	20.1%
Ordered	28,512	26,887	-5.7%

$p = 2.713*10^{-93}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	333	191	-42.6%
Ordered	1,222	1,364	11.6%

$p = 1.664*10^{-18}$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	401	191	-52.4%
Ordered	1,154	1,364	18.2%

$p = 5.097*10^{-34}$

CGP_br/col_1

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	967	1,116	15.4%
Ordered	2,569	2,420	-5.8%

$p = 1.894*10^{-8}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	317	313	-1.3%
Ordered	922	926	0.4%

$p = 0.7945$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	367	313	-14.7%
Ordered	872	926	6.2%

$p = 7.437*10^{-4}$

CGP_CAN_br/col_1

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	245	277	13.1%
Ordered	663	631	-4.8%

$p = 1.673*10^{-2}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	109	97	-11.0%
Ordered	286	298	4.2%

$p = 0.1768$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	123	97	-21.1%
Ordered	272	298	9.6%

$p = 4.575*10^{-3}$

CGP_br/col_2

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	1,772	2,036	14.9%
Ordered	4,326	4,062	-6.1%

$p =9.622*10^{-14}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	448	465	3.8%
Ordered	1,291	1,274	-1.3%

$p = 0.3513$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	518	465	-10.2%
Ordered	1,221	1,274	4.3%

$p = 5.451*10^{-3}$

CGP_CAN_br/col_2

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	470	506	7.7%
Ordered	902	866	-4.0%

$p = 4.056*10^{-2}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	151	131	-13.2%
Ordered	362	382	5.5%

$p = 5.268*10^{-2}$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	163	131	-19.6%
Ordered	350	382	9.1%

$p = 2.410*10^{-3}$

CGP_pan

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	1,090	1,206	10.6%
Ordered	2,758	2,642	-4.2%
$p = 3.322*10^{-5}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	197	205	4.1%
Ordered	593	585	-1.3%
p = 0.5106			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	219	205	-6.4%
Ordered	571	585	2.5%
p = 0.2658			

CGP_CAN_pan

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	78	84	7.7%
Ordered	211	205	-2.8%
p = 0.4266			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	36	32	-11.1%
Ordered	94	98	4.3%
p = 0.4330			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	39	32	-17.9%
Ordered	91	98	7.7%
p = 0.1803			

CGP_glio

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	1,626	1,898	16.7%
Ordered	4,168	3,896	-6.5%
$p = 1.820*10^{-15}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	274	291	6.2%
Ordered	921	904	-1.8%
p = 0.2421			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	324	291	-10.2%
Ordered	871	904	3.8%
$p = 3.329*10^{-2}$			

CGP_CAN_glio

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	37	54	45.9%
Ordered	173	156	-9.8%
$p = 2.076*10^{-3}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	21	16	-23.8%
Ordered	56	61	8.9%
p = 0.2008			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	29	16	-44.8%
Ordered	48	61	27.1%
$p = 2.103*10^{-3}$			

COSMIC

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	6,568	7,784	18.5%
Ordered	19,867	18,651	-6.1%
$p = 4.112*10^{-67}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	5,781	5,215	-9.8%
Ordered	16,927	17,493	3.3%
$p = 6.572*10^{-18}$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	6,841	5,215	-23.8%
Ordered	15,867	17,493	10.2%
$p = 2.679*10^{-122}$			

COSMIC_census

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	253	303	19.8%
Ordered	420	370	-11.9%
$p = 6.916*10^{-5}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered	1,633	1,125	-31.1%
Ordered	3,742	4,250	13.6%
$p = 2.696*10^{-51}$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered	2,001	1,125	-43.8%
Ordered	3,374	4,250	26.0%
$p = 5.316*10^{-135}$			

Table S3

SP_cancer

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	4470	4870	8.95%
Disordered	5051	6123	21.22%
Ordered	27062	25590	-5.44%

$p = 2.529*10^{-75}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	1032	804	-22.09%
Disordered	988	842	-14.78%
Ordered	3226	3600	11.59%

$p = 8.000*10^{-26}$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	1098	804	-26.78%
Disordered	1170	842	-28.03%
Ordered	2978	3600	20.89%

$p = 5.349*10^{-66}$

SP_cancer_annotated

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	4470	4870	8.95%
Disordered	5051	6123	21.22%
Ordered	27062	25590	-5.44%

$p = 2.529*10^{-75}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	187	144	-22.99%
Disordered	226	146	-35.40%
Ordered	1142	1265	10.77%

$p = 7.029*10^{-12}$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	203	144	-29.06%
Disordered	274	146	-46.72%
Ordered	1078	1265	17.35%

$p = 1.770*10^{-24}$

CGP_br/col_1

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	581	623	7.23%
Disordered	551	643	16.70%
Ordered	2404	2270	-5.57%

$p = 2.416*10^{-6}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	185	173	-6.49%
Disordered	190	195	2.63%
Ordered	864	871	0.81%

$p = 0.6167$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	199	173	-13.07%
Disordered	222	195	-12.16%
Ordered	818	871	6.48%

$p = 6.362*10^{-3}$

CGP_CAN_br/col_1

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	152	156	2.63%
Disordered	130	153	17.69%
Ordered	626	599	-4.31%

$p = 6.929*10^{-2}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	67	64	-4.48%
Disordered	61	56	-8.20%
Ordered	267	275	3.00%

$p = 0.6758$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	68	64	-5.88%
Disordered	72	56	-22.22%
Ordered	255	275	7.84%

$p = 6.858*10^{-2}$

CGP_br/col_2

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	1056	1118	5.87%
Disordered	1011	1173	16.02%
Ordered	4031	3807	-5.56%

$p = 7.410*10^{-10}$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	262	249	-4.96%
Disordered	268	292	8.96%
Ordered	1209	1198	-0.91%

$p = 0.2352$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	279	249	-10.75%
Disordered	312	292	-6.41%
Ordered	1148	1198	4.36%

$p = 3.534*10^{-2}$

CGP_CAN_br/col_2

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	309	322	4.21%
Disordered	229	245	6.99%
Ordered	834	805	-3.48%

$p = 0.2627$

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	95	90	-5.26%
Disordered	82	75	-8.54%
Ordered	336	348	3.57%

$p = 0.5249$

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	99	90	-9.09%
Disordered	88	75	-14.77%
Ordered	326	348	6.75%

$p = 0.1210$

CGP_pan

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	634	631	-0.47%
Disordered	631	720	14.10%
Ordered	2583	2497	-3.33%
$p = 4.460*10^{-4}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	113	118	4.42%
Disordered	121	128	5.79%
Ordered	556	544	-2.16%
$p = 0.6424$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	113	118	4.42%
Disordered	138	128	-7.25%
Ordered	539	544	0.93%
$p = 0.6089$			

CGP_CAN_pan

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	45	53	17.78%
Disordered	48	47	-2.08%
Ordered	196	189	-3.57%
$p = 0.4289$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	23	24	4.35%
Disordered	22	18	-18.18%
Ordered	85	88	3.53%
$p = 0.6451$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	27	24	-11.11%
Disordered	21	18	-14.29%
Ordered	82	88	7.32%
$p = 0.5486$			

CGP_glio

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	972	1043	7.30%
Disordered	908	1082	19.16%
Ordered	3914	3669	-6.26%
$p = 2.010*10^{-12}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	158	154	-2.53%
Disordered	164	185	12.80%
Ordered	873	856	-1.95%
$p = 0.2100$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	171	154	-9.94%
Disordered	197	185	-6.09%
Ordered	827	856	3.51%
$p = 0.1793$			

CGP_CAN_glio

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	24	33	37.50%
Disordered	19	27	42.11%
Ordered	167	150	-10.18%
$p=1.445*10^{-2}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	13	12	-7.69%
Disordered	11	8	-27.27%
Ordered	53	57	7.55%
$p = 0.5496$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	17	12	-29.41%
Disordered	15	8	-46.67%
Ordered	45	57	26.67%
$p = 1.890*10^{-2}$			

COSMIC

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	3800	4097	7.82%
Disordered	3915	4750	21.33%
Ordered	18720	17588	-6.05%
$p = 2.650*10^{-59}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	3383	3037	-10.23%
Disordered	3551	3362	-5.32%
Ordered	15774	16309	3.39%
$p = 1.553*10^{-14}$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	3637	3037	-16.50%
Disordered	4295	3362	-21.72%
Ordered	14776	16309	10.37%
$p = 9.101*10^{-101}$			

COSMIC_census

Polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	155	157	1.29%
Disordered	137	174	27.01%
Ordered	381	342	-10.24%
$p = 9.071*10^{-4}$			

Cancer mutations

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	1033	807	-21.88%
Disordered	992	760	-23.39%
Ordered	3350	3808	13.67%
$p = 7.663*10^{-37}$			

Cancer mutations compared to polymorphisms

	Expected	Observed	Over-representation compared to expected
Disordered binding sites	1058	807	-23.72%
Disordered	1275	760	-40.39%
Ordered	3042	3808	25.18%
$p = 1.033*10^{-100}$			

Table S4

	Bin 1 boundaries	**Bin 2 boundaries**	**Bin 3 boundaries**	**Bin 4 boundaries**	**Bin 5 boundaries**
Length	<200	200-400	400-600	600-900	900+
Disorder %	<20%	20-40%	40-60%	60-80%	80%+
Binding regions %	<5%	5-15%	15-25%	25-40%	40%+
COSMIC census mutations	0	1	2	3-4	5+
Interactions	0	1-5	6-10	11-50	51+