

Supplementary information S2 | **Number of proteins in the cancer set with Pfam domains**

Pfam domain	Number of cancer proteins with domain	Number of human proteins with domain	Expected number of proteins with domain in a randomly chosen sample of human proteins	Significant (at 5%)
Pkinase	27	643	6.3	S
HLH	12	123	1.2	S
Homeobox	13	231	2.3	S
Ets	6	38	0.4	S
PAX	4	11	0.1	S
PHD	8	100	1.0	S
DNA_mis_repair	3	4	0.0	S
HATPase_c	5	28	0.3	S
Myc_N_term	3	5	0.0	S
AF-4	3	7	0.1	S
Bromodomain	5	49	0.5	S
DUF902	2	2	0.0	NS
KIX	2	2	0.0	NS
SH2	7	129	1.3	NS
PWWP	4	33	0.3	NS
DUF906	2	3	0.0	NS
HRDC	2	3	0.0	NS
zf-TAZ	2	3	0.0	NS
ETS_PEA3_N	2	3	0.0	NS
SAM_PNT	3	17	0.2	NS
MutS_I	2	4	0.0	NS
Exostosin	2	4	0.0	NS
SSXT	2	5	0.0	NS
Nucleoporin_FG	2	5	0.0	NS
MutS_II	2	5	0.0	NS
SH3	9	262	2.6	NS
GTP_CDC	3	22	0.2	NS
DEAD	5	83	0.8	NS
ERM	2	6	0.1	NS
YEATS	2	6	0.1	NS
HSP90	2	7	0.1	NS
MutS_IV	2	7	0.1	NS
AT_hook	3	26	0.3	NS
Helicase_C	6	137	1.3	NS
MutS_III	2	8	0.1	NS
Recep_L_domain	2	8	0.1	NS
Furin-like	2	8	0.1	NS
zf-RanBP	3	29	0.3	NS
7tm_1	1	800	7.8	S
MOZ_SAS	2	9	0.1	NS
Tropomyosin	2	9	0.1	NS
MutS_V	2	11	0.1	NS
ENTH	2	13	0.1	NS
Myosin_N	2	13	0.1	NS
Armadillo_seg	3	42	0.4	NS
COX6C	1	1	0.0	NS
AICARFT_IMPCHas	1	1	0.0	NS
Ribosomal_L22e	1	1	0.0	NS
Tuberin	1	1	0.0	NS
Fanconia	1	1	0.0	NS
APC_15aa	1	1	0.0	NS
GMP_synt_C	1	1	0.0	NS

CybS	1	1	0.0	NS
Hamartin	1	1	0.0	NS
EB1_binding	1	1	0.0	NS
XPA_N	1	1	0.0	NS
Cdc73	1	1	0.0	NS
WT1	1	1	0.0	NS
HNF-1A_C	1	1	0.0	NS
VHL	1	1	0.0	NS
Myc-LZ	1	1	0.0	NS
UPF0023	1	1	0.0	NS
BRCA2	1	1	0.0	NS
XPA_C	1	1	0.0	NS
Fanconi	1	1	0.0	NS
Fork_head	3	49	0.5	NS
LIM	4	92	0.9	NS
Myosin_tail	2	19	0.2	NS
CARD	2	21	0.2	NS
Menin	1	2	0.0	NS
CBF_beta	1	2	0.0	NS
APC_basic	1	2	0.0	NS
HNF-1B_C	1	2	0.0	NS
Topoisomerase_I	1	2	0.0	NS
Topoisomer_I_N	1	2	0.0	NS
PDGF_N	1	2	0.0	NS
Nucleoplasmin	1	2	0.0	NS
SUFU	1	2	0.0	NS
MoeA_N	1	2	0.0	NS
SNF5	1	2	0.0	NS
Fip1	1	2	0.0	NS
MoeA_C	1	2	0.0	NS
Clathrin_propel	1	2	0.0	NS
Sdh_cyt	1	2	0.0	NS
Rad4	1	2	0.0	NS
zf-MYND	2	24	0.2	NS
hormone_rec	3	62	0.6	NS
zf-C4	3	63	0.6	NS
ZZ	2	25	0.2	NS
SAMP	1	3	0.0	NS
TCL1_MTCP1	1	3	0.0	NS
RB_A	1	3	0.0	NS
Caudal_act	1	3	0.0	NS
RB_B	1	3	0.0	NS
Rabaptin	1	3	0.0	NS
Nucleoporin2	1	3	0.0	NS
Runt	1	3	0.0	NS
MGS	1	3	0.0	NS
APC_crr	1	3	0.0	NS
BH4	1	3	0.0	NS
ERCC4	1	3	0.0	NS
YLP	1	3	0.0	NS
FHA	2	27	0.3	NS
Cbl_N2	1	4	0.0	NS
Cbl_N3	1	4	0.0	NS
lyase_1	1	4	0.0	NS
Cbl_N	1	4	0.0	NS

HNF-1_N	1	4	0.0	NS
Band_41	2	33	0.3	NS
XPG_I	1	5	0.0	NS
Hox9_act	1	5	0.0	NS
P53	1	5	0.0	NS
TIG	2	37	0.4	NS
RhoGAP	3	85	0.8	NS
death	2	39	0.4	NS
FYRN	1	6	0.1	NS
MoCF_biosynth	1	6	0.1	NS
FYRC	1	6	0.1	NS
HhH-GPD	1	6	0.1	NS
pKID	1	6	0.1	NS
I_LWEQ	1	6	0.1	NS
Glypican	1	6	0.1	NS
notch	1	6	0.1	NS
GATase	1	6	0.1	NS
PBX	1	6	0.1	NS
TFR_dimer	1	6	0.1	NS
BCL_N	1	6	0.1	NS
PAS	2	40	0.4	NS
zf-B_box	3	93	0.9	NS
fer2	1	7	0.1	NS
XPG_N	1	7	0.1	NS
Patched	1	7	0.1	NS
RBD	1	7	0.1	NS
FAT	1	7	0.1	NS
Clathrin	1	7	0.1	NS
ras	4	155	1.5	NS
3_5_exonuclease	1	8	0.1	NS
DIL	1	8	0.1	NS
MH2	1	8	0.1	NS
BTK	1	8	0.1	NS
Anti_proliferat	1	8	0.1	NS
myosin_head	2	48	0.5	NS
NAC	1	9	0.1	NS
HHH	1	9	0.1	NS
FATC	1	9	0.1	NS
Peptidase_M28	1	9	0.1	NS
WH1	1	9	0.1	NS
Nebulin	1	9	0.1	NS
UBA	2	51	0.5	NS
SET	2	52	0.5	NS
IRF	1	10	0.1	NS
PDGF	1	10	0.1	NS
Rlla	1	11	0.1	NS
BIR	1	11	0.1	NS
RHD	1	11	0.1	NS
Activin_recp	1	12	0.1	NS
zf-C3HC4	5	262	2.6	NS
cyclin_C	1	13	0.1	NS
Frizzled	1	14	0.1	NS
OAR	1	14	0.1	NS
MH1	1	14	0.1	NS
Peptidase_C14	1	14	0.1	NS

Rap_GAP	1	14	0.1	NS
CAP_GLY	1	14	0.1	NS
ig	11	763	7.5	NS
RasGAP	1	15	0.1	NS
NAP	1	15	0.1	NS
GATA	1	15	0.1	NS
WD40	1	344	3.4	NS
COLFI	1	17	0.2	NS
DAG_PE-bind	2	72	0.7	NS
zf-CXXC	1	18	0.2	NS
BTB	3	146	1.4	NS
Bcl-2	1	20	0.2	NS
PB1	1	20	0.2	NS
UBX	1	20	0.2	NS
MAM	1	20	0.2	NS
NACHT	1	20	0.2	NS
WH2	1	22	0.2	NS
PA	1	22	0.2	NS
RhoGEF	2	84	0.8	NS
FCH	1	23	0.2	NS
LRRCT	2	87	0.9	NS
NUDIX	1	24	0.2	NS
Cadherin_C_term	1	24	0.2	NS
TNFR_c6	1	25	0.2	NS
PBD	1	25	0.2	NS
LRR	4	244	2.4	NS
PSI	1	26	0.3	NS
PI3_PI4_kinase	1	26	0.3	NS
G-alpha	1	27	0.3	NS
BRCT	1	27	0.3	NS
Fz	1	28	0.3	NS
SAP	1	29	0.3	NS
IQ	2	110	1.1	NS
cyclin	1	37	0.4	NS
cNMP_binding	1	37	0.4	NS
AMP-binding	1	38	0.4	NS
rrm	4	290	2.8	NS
kazal	1	39	0.4	NS
Sema	1	40	0.4	NS
zf-C2H2	10	871	8.5	NS
RA	1	44	0.4	NS
pkinase_C	1	45	0.4	NS
EGF	1	219	2.1	NS
histone	1	49	0.5	NS
bZIP	1	51	0.5	NS
LRRNT	1	56	0.5	NS
Y_phosphatase	1	59	0.6	NS
cadherin	2	162	1.6	NS
ank	3	266	2.6	NS
HMG_box	1	68	0.7	NS
WW	1	72	0.7	NS
UCH	1	72	0.7	NS
CH	1	78	0.8	NS
PDZ	2	186	1.8	NS
PH	3	303	3.0	NS

C2	1	146	1.4	NS
Collagen	1	90	0.9	NS
efhand	2	233	2.3	NS
KRAB	3	308	3.0	NS
All proteins	284	29064		

The results shown are based on 284 proteins in the cancer set for which an amino-acid sequence could be obtained. These proteins have been compared both with the number of proteins in the human proteome with a particular Pfam domain, and with the expected number of proteins with the domain in a randomly chosen sample of 284 human proteins. The non-redundant human Swiss-Prot/TrEMBL proteome set (29,064 proteins) from the European Bioinformatics Institute was used as the reference human proteome set (<http://www.ebi.ac.uk/proteome/>). S, probabilities that are significant at 5% significance; NS, probabilities that are not significant.