

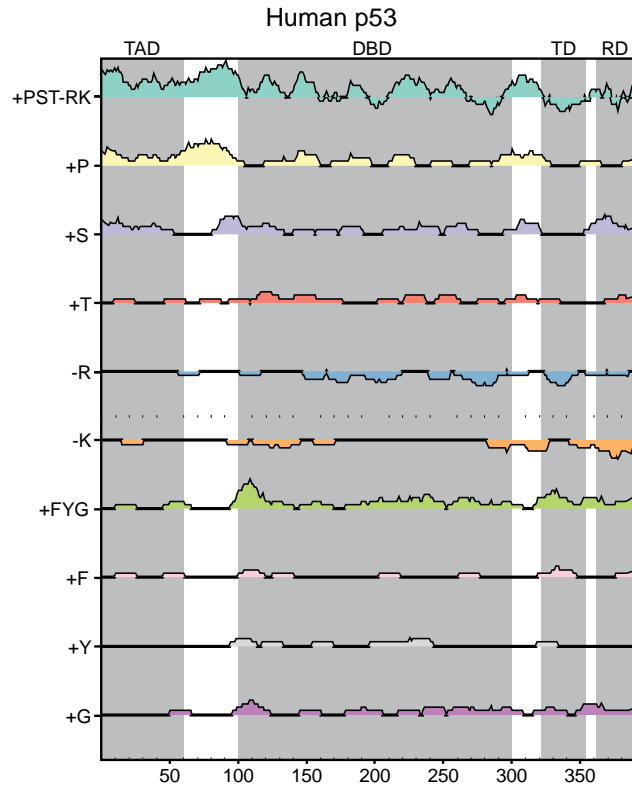
Supporting information

IDDomainSpotter: Compositional bias reveals domains in long disordered protein regions - insights from transcription factors

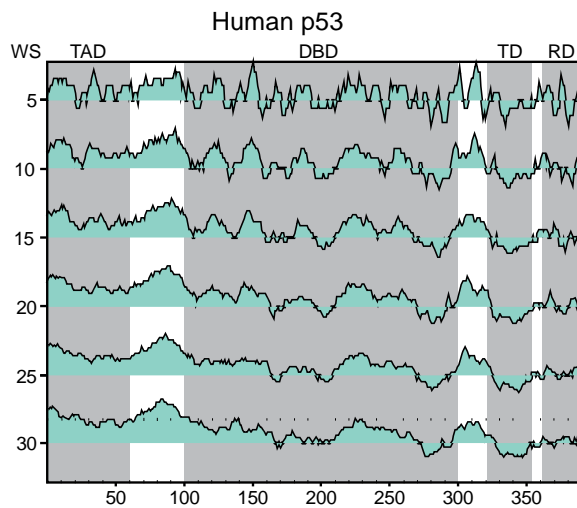
Peter S. Millard^{1,2}, Katrine Bugge³, Riccardo Marabini³, Wouter K. Boomsma^{4*}, Meike Burow^{1,2}, Birthe B. Kragelund^{3*}

¹DynaMo Center, Dept. of Plant and Environmental Sciences, University of Copenhagen, Denmark, ²Copenhagen Plant Science Centre, Dept. of Plant and Environmental Sciences, University of Copenhagen, Denmark. ³Structural Biology and NMR Laboratory, Dept. of Biology, University of Copenhagen, Denmark. ⁴Department of Computer Science, University of Copenhagen, Denmark.

*Corresponding author (Birthe B. Kragelund, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark, +45 3532 2081, bbk@bio.ku.dk) (Correspondence may also be addressed to Wouter Boomsma, wb@di.ku.dk)



Supplementary Figure 1. IDDomainSpotter profiles of human p53. Grey boxes indicate domains, conserved regions and motifs as named above the graph. DBD: DNA-binding domain, TAD: transactivation domain, TD: tetramerization domain, RD: regulatory domain. Profiles display scores for Pro+Ser+Thr-Arg-Lys (+PST-RK), Phe+Tyr+Gly (+FYG) and individual residues, calculated over 15 residues windows.



Supplementary Figure 2. IDDomainSpotter profiles of human p53. Grey boxes indicate domains, conserved regions and motifs as named above the graph. DBD: DNA-binding domain, TAD: transactivation domain, TD: tetramerization domain, RD: regulatory domain. Profiles display scores for Pro+Ser+Thr-Arg-Lys (+PST-RK), calculated with various window sizes (WS).

Supplementary table 1. Sequences used in Mammalia p53 alignment (Figure 1C).

UniProt identifier	Species	Species common name
P41685	<i>Felis catus</i>	Cat
A0A3Q7Y5W2	<i>Ursus arctos horribilis</i>	Grizzly bear
G1MEP6	<i>Ailuropoda melanoleuca</i>	Giant panda
A0A2U3VT95	<i>Odobenus rosmarus divergens</i>	Pacific walrus
M3YC88	<i>Mustela putorius furo</i>	European domestic ferret
A0A3Q7TIN0	<i>Vulpes vulpes</i>	Red fox
F1PI27	<i>Canis lupus familiaris</i>	Dog
A0A2U3Y5G3	<i>Leptonychotes weddellii</i>	Weddell seal
A0A2Y9L4Q2	<i>Enhydra lutris kenyon</i>	Sea otter
A0A455C1G1	<i>Physeter macrocephalus</i>	Sperm whale
A0A2U4C2U9	<i>Tursiops truncatus</i>	Atlantic bottle-nosed dolphin
A0A340XCN5	<i>Lipotes vexillifer</i>	Yangtze river dolphin
A0A2Y9Q793	<i>Delphinapterus leucas</i>	Beluga whale
A0A383YUA7	<i>Balaenoptera acutorostrata scammoni</i>	North Pacific minke whale
G3T035	<i>Loxodonta africana</i>	African elephant
A0A2Y9R5B9	<i>Trichechus manatus latirostris</i>	Florida manatee
A0A1S2ZRH2	<i>Erinaceus europaeus</i>	Western European hedgehog
C3VC56	<i>Ovis aries</i>	Sheep
P67939	<i>Bos taurus</i>	Bovine
A0A3Q2HE67	<i>Equus caballus</i>	Horse
L7N1B2	<i>Myotis lucifugus</i>	Little brown bat
L5JZ91	<i>Pteropus alecto</i>	Black flying fox
Q9TUB2	<i>Sus scrofa</i>	Pig
Q95330	<i>Oryctolagus cuniculus</i>	Rabbit
G3WS63	<i>Sarcophilus harrisii</i>	Tasmanian devil
A0A158SIS7	<i>Mus musculus</i>	Mouse
P10361	<i>Rattus norvegicus</i>	Rat
A0A1S3EYC1	<i>Dipodomys ordii</i>	Ord's kangaroo rat
A0A1U7Q2G4	<i>Mesocricetus auratus</i>	Golden hamster
Q9WUR6	<i>Cavia porcellus</i>	Guinea pig
I3N5N2	<i>Ictidomys tridecemlineatus</i>	Thirteen-lined ground squirrel
G5B5D6	<i>Heterocephalus glaber</i>	Naked mole rat
A0A091CJU7	<i>Fukomys damarensis</i>	Damaraland mole rat

Supplementary table 2. Sequences used in Chordata p53 alignment (Figure 1E).

UniProt identifier	Species	Species common name
C0H8X1	<i>Salmo salar</i>	Atlantic salmon
A0A060XN97	<i>Oncorhynchus mykiss</i>	Rainbow trout
A0A3F2YPG3	<i>Xiphophorus maculatus</i>	Southern platyfish
A0A2R8RHX5	<i>Danio rerio</i>	Zebrafish
W5LNF4	<i>Astyanax mexicanus</i>	Blind cave fish
O93379	<i>Ictalurus punctatus</i>	Channel catfish
H2U133	<i>Takifugu rubripes</i>	Japanese pufferfish
H3CXQ0	<i>Tetraodon nigroviridis</i>	Spotted green pufferfish
W5N8V4	<i>Lepisosteus oculatus</i>	Spotted gar
G3PK82	<i>Gasterosteus aculeatus</i>	Three-spined stickleback
I3KRX9	<i>Oreochromis niloticus</i>	Nile tilapia
A0A2U9B0Y2	<i>Scophthalmus maximus</i>	Turbot
P10360	<i>Gallus gallus</i>	Chicken
A0A2I0M383	<i>Columba livia</i>	Rock dove
A0A493TL36	<i>Anas platyrhynchos platyrhynchos</i>	Northern mallard
A0A1U7SJ11	<i>Alligator sinensis</i>	Chinese alligator
A0A151MW63	<i>Alligator mississippiensis</i>	American alligator
F7A9M9	<i>Xenopus tropicalis</i>	Western clawed frog
H2ZH75	<i>Ciona savignyi</i>	Pacific transparent sea squirt
Q4H300	<i>Ciona intestinalis</i>	Transparent sea squirt
H3B1Z4	<i>Latimeria chalumnae</i>	Coelacanth
K7G3P4	<i>Pelodiscus sinensis</i>	Chinese softshell turtle

Supplementary table 3. Sequences used in Mammalia GCR alignment (Figure 1D).

UniProt identifier	Species	Species common name
M3WZ39	<i>Felis catus</i>	Cat
A0A3Q7W0S9	<i>Ursus arctos horribilis</i>	Grizzly bear
A0A2U3WVZ7	<i>Odobenus rosmarus divergens</i>	Pacific walrus
A0A2Y9KIM9	<i>Enhydra lutris kenyoni</i>	Sea otter
A0A3Q7V3D8	<i>Vulpes vulpes</i>	Red fox
A0A2U3XTM7	<i>Leptonychotes weddellii</i>	Weddell seal
A0A384C5F8	<i>Ursus maritimus</i>	Polar bear
A0A452SBP1	<i>Ursus americanus</i>	American black bear
F1Q298	<i>Canis lupus familiaris</i>	Dog
G1LUC3	<i>Ailuropoda melanoleuca</i>	Giant panda
M3Y916	<i>Mustela putorius furo</i>	European domestic ferret
A0A2Y9PIT4	<i>Delphinapterus leucas</i>	Beluga whale
A0A2Y9SPF1	<i>Physeter macrocephalus</i>	Sperm whale
A0A384B3U4	<i>Balaenoptera acutorostrata scammoni</i>	North Pacific minke whale
A0A340Y3E1	<i>Lipotes vexillifer</i>	Yangtze river dolphin
A0A2U4AH03	<i>Tursiops truncatus</i>	Atlantic bottle-nosed dolphin
F1MN19	<i>Bos taurus</i>	Bovine
A0A452EXA0	<i>Capra hircus</i>	Goat
W5NVF6	<i>Ovis aries</i>	Sheep
F1RM35	<i>Sus scrofa</i>	Pig
A0A2Y9DU07	<i>Trichechus manatus latirostri</i>	Florida manatee
A0A1S2ZGL3	<i>Erinaceus europaeus</i>	Western European hedgehog
G1PHX2	<i>Myotis lucifugus</i>	Little brown bat
A0A3Q2HH04	<i>Equus caballus</i>	Horse
G3SVD4	<i>Loxodonta africana</i>	African elephant
G3VFI8	<i>Sarcophilus harrisii</i>	Tasmanian devil
G1T1R4	<i>Oryctolagus cuniculus</i>	Rabbit
F6S9Z9	<i>Monodelphis domestica</i>	Gray short-tailed opossum
F6W567	<i>Ornithorhynchus anatinus</i>	Duckbill platypus
L9L2B3	<i>Tupaia chinensis</i>	Chinese tree shrew
P06537	<i>Mus musculus</i>	Mouse
P06536	<i>Rattus norvegicus</i>	Rat
A0A1U8BLX0	<i>Mesocricetus auratus</i>	Golden hamster
H0V4W0	<i>Cavia porcellus</i>	Guinea pig
A0A1S3FH64	<i>Dipodomys ordii</i>	Ord's kangaroo rat
A0A287D1I2	<i>Ictidomys tridecemlineatus</i>	Thirteen-lined ground squirrel
G5AQS2	<i>Heterocephalus glaber</i>	Naked mole rat
G3HEK6	<i>Cricetulus griseus</i>	Chinese hamster
A0A091DN95	<i>Fukomys damarensis</i>	Damaraland mole rat

Supplementary table 4. Sequences used in Chordata GCR alignment (Figure 1F).

UniProt identifier	Species	Species common name
A0A1S3LMP5	<i>Salmo salar</i>	Atlantic salmon
A0A3B4AXU8	<i>Periophthalmus magnuspinnatus</i>	
A0A3P8Z8R7	<i>Esox lucius</i>	Northern pike
A0A3B3QRS8	<i>Paramormyrops kingsleyae</i>	
Q1XHK0	<i>Danio rerio</i>	Zebrafish
A0A3B5A6Z0	<i>Stegastes partitus</i>	Bicolor damselfish
A0A2D0SEE8	<i>Ictalurus punctatus</i>	Channel catfish
A0A3B3HHZ3	<i>Oryzias latipes</i>	Japanese rice fish
A0A3B4XM90	<i>Seriola lalandi dorsalis</i>	California yellowtail
A0A3B4TRJ1	<i>Seriola dumerili</i>	Greater amberjack
A0A3P9C0Z8	<i>Maylandia zebra</i>	Zebra mbuna
A0A3P8NYX9	<i>Astatotilapia calliptera</i>	Eastern happy
A0A3B3YXW0	<i>Poecilia mexicana</i>	Atlantic molly
A0A3B4D9J0	<i>Pygocentrus nattereri</i>	Red-bellied piranha
A0A3B1IJY8	<i>Astyanax mexicanus</i>	Blind cave fish
A0A3B4ET53	<i>Pundamilia nyererei</i>	
A0A2I4BX66	<i>Austrofundulus limnaeus</i>	
A0A060WHU5	<i>Oncorhynchus mykiss</i>	Rainbow trout
A0A3B3BLV6	<i>Oryzias melastigma</i>	Marine medaka
M4AXU0	<i>Xiphophorus maculatus</i>	Southern platyfish
A0A3P8TUC7	<i>Amphiprion percula</i>	Orange clownfish
A0A3P8WGG5	<i>Cynoglossus semilaevis</i>	Tongue sole
A0A3B3V5Z8	<i>Poecilia latipinna</i>	Sailfin molly
H2T175	<i>Takifugu rubripes</i>	Japanese pufferfish
I3KN75	<i>Oreochromis niloticus</i>	Nile tilapia
A0A087XJA9	<i>Poecilia formosa</i>	Amazon molly
A0A3P9NF67	<i>Poecilia reticulata</i>	Guppy
G3Q2H5	<i>Gasterosteus aculeatus</i>	Three-spined stickleback
H3CU66	<i>Tetraodon nigroviridis</i>	Spotted green pufferfish
A0A2U9C6P4	<i>Scophthalmus maximus</i>	Turbot
W5MWX4	<i>Lepisosteus oculatus</i>	Spotted gar
A0A1D5PRD7	<i>Gallus gallus</i>	Chicken
A0A493TR65	<i>Anas platyrhynchos platyrhynchos</i>	Northern mallard
A0A2I0LTB5	<i>Columba livia</i>	Rock dove
A0A1V4KTY0	<i>Patagioenas fasciata monilis</i>	
H0YPE9	<i>Taeniopygia guttata</i>	Zebra finch
A0A3M0J6J1	<i>Hirundo rustica rustica</i>	Barn swallow
A0A226MKI6	<i>Callipepla squamata</i>	Scaled quail
A0A218UEU6	<i>Lonchura striata domestica</i>	Bengalese finch
U3K2U0	<i>Ficedula albicollis</i>	Collared flycatcher
G1NAS4	<i>Meleagris gallopavo</i>	Wild turkey
A0A0Q3P7Y5	<i>Amazona aestiva</i>	Blue-fronted Amazon parrot
A0A091W3R1	<i>Nipponia nippon</i>	Crested ibis
A0A091FM40	<i>Cuculus canorus</i>	Common cuckoo
A0A091IR26	<i>Calypte anna</i>	Anna's hummingbird
A0A087RF69	<i>Aptenodytes forsteri</i>	Emperor penguin
A0A091F184	<i>Corvus brachyrhynchos</i>	American crow
A0A099ZXG0	<i>Charadrius vociferus</i>	Killdeer
A0A099YVI9	<i>Tinamus guttatus</i>	White-throated tinamou
A0A093GET9	<i>Dryobates pubescens</i>	Downy woodpecker
A0A091IP29	<i>Egretta garzetta</i>	Little egret
A0A091W088	<i>Opisthocomus hoazin</i>	Hoatzin
A0A093PLE0	<i>Manacus vitellinus</i>	Golden-collared manakin

A0A093H389	<i>Struthio camelus australis</i>	South African ostrich
A0A452HUM1	<i>Gopherus agassizii</i>	Agassiz's desert tortoise
K7GAE8	<i>Pelodiscus sinensis</i>	Chinese softshell turtle
A0A3Q0GJF3	<i>Alligator sinensis</i>	Chinese alligator
A0A151M6S0	<i>Alligator mississippiensis</i>	American alligator
Q6DDL0	<i>Xenopus laevis</i>	African clawed frog
F6XE59	<i>Xenopus tropicalis</i>	Western clawed frog
G1KP59	<i>Anolis carolinensis</i>	Green anole
A0A401T225	<i>Chiloscyllium punctatum</i>	Brownbanded bambooshark
A0A401P986	<i>Scyliorhinus torazame</i>	Cloudy catshark
H3B6W7	<i>Latimeria chalumnae</i>	Coelacanth

Supplementary table 5. Sequences used in Brassicales NAC46 alignment (Figure 2C).

UniProt identifier	Species	Species common name
A0A078IN86	<i>Brassica napus</i>	Rape
A0A0D3AFP5	<i>Brassica oleracea var. oleracea</i>	Wild cabbage
M4ET20	<i>Brassica rapa subsp. pekinensis</i>	Chinese cabbage
R0G5Z5	<i>Capsella rubella</i>	Pink shepherd's-purse
V4LGP0	<i>Eutrema salsugineum</i>	Saltwater cress

Supplementary table 6. Sequences used in Brassicales NAC13 alignment (Figure 2D).

UniProt identifier	Species	Species common name
A0A078FDS7	<i>Brassica napus</i>	Rape
A0A0D3D7A3	<i>Brassica oleracea var. oleracea</i>	Wild cabbage
M4D782	<i>Brassica rapa subsp. pekinensis</i>	Chinese cabbage
R0IN97	<i>Capsella rubella</i>	Pink shepherd's-purse
V4MRY1	<i>Eutrema salsugineum</i>	Saltwater cress

Supplementary table 7. Sequences used in Brassicales MYB28/MYB29-like alignment (Figure 3C).

UniProt identifier	Species	Species common name
A0A0D3DDK8	<i>Brassica oleracea var. oleracea</i>	Wild cabbage
A0A078JG35	<i>Brassica napus</i>	Rape
M4D901	<i>Brassica rapa subsp. pekinensis</i>	Chinese cabbage
R0GD41	<i>Capsella rubella</i>	Pink shepherd's-purse
V4KWQ2	<i>Eutrema salsugineum</i>	Saltwater cress
A0A0D3EH73	<i>Brassica oleracea var. oleracea</i>	Wild cabbage
C5HYG2	<i>Brassica rapa subsp. pekinensis</i>	Chinese cabbage
R0GVT9	<i>Capsella rubella</i>	Pink shepherd's-purse
V4LLL3	<i>Eutrema salsugineum</i>	Saltwater cress