Supplemental Materials

Molecular Biology of the Cell

Hutchins

Supplemental Materials for:

What's that gene (or protein)?

Online resources for exploring functions of genes, transcripts and proteins

James R. A. Hutchins

Institute of Human Genetics, CNRS, 141 rue de la Cardonille, 34396 Montpellier, France.

Email: james.hutchins@igh.cnrs.fr

Note:

Periodic updates to this Supplemental Materials document will be made available on the author's personal website: http://www.jrahutchins.net

Supplemental Table S1

Species-specific gene databases

Species	Resource	Example web-links	Reference
Arabidopsis thaliana	TAIR	Home page: http://www.arabidopsis.org/ Query via (a) gene name (b) locus (c) UniProt ID: (a) http://www.arabidopsis.org/servlets/Search?type=general&search_action=detail&sub_type=gene&name=cdc2 (b) http://www.arabidopsis.org/servlets/Search?type=general&search_action=detail&sub_type=gene&name=AT1G08260 (DNA pol) (c) http://www.arabidopsis.org/servlets/Search?type=general&search_action=detail&sub_type=protein&name=F4KFS5	(Lamesch *et al.*, 2012)
Ascidians (various species)	ANISEED	Home page: http://www.aniseed.cnrs.fr/ Query via any term, e.g. gene name: http://www.aniseed.cnrs.fr/aniseed/anisearch/?search=Cdc2 (Cdk1) Query via species name plus gene term: http://www.aniseed.cnrs.fr/v3/gene-result.php?species=Ciona+intestinalis&name=Pole (DNA pol) Direct link to Transcript model card via Ensembl transcript ID: http://www.aniseed.cnrs.fr/v3/molecule-gene.php?name=ENSCINT00000009154	(Tassy *et al.*, 2010)
Aspergillus (various species)	AspGD	Home page: http://www.aspergillusgenome.org/ Direct link to Summary page via (a) gene name (b) systematic name (c) AspGD ID: (a) http://www.aspergillusgenome.org/cgi-bin/locus.pl?locus=nimX (Cdk1) (b) http://www.aspergillusgenome.org/cgi-bin/locus.pl?locus=AN3067 (DNA pol) (c) http://www.aspergillusgenome.org/cgi-bin/locus.pl?locus=ASPL0000080189 Query via UniProt ID: http://www.aspergillusgenome.org/cgi-bin/search/textSearch?query=P11837	(Arnaud *et al.*, 2012)
Caenorhabditis elegans	WormBase	Home page: http://www.wormbase.org/ Direct link to Gene page via (a) gene name (b) sequence code (c) WormBase ID (d) RefSeq nucleotide: (a) http://www.wormbase.org/search/gene/cdk-1 (Cdk1) (b) http://www.wormbase.org/search/gene/F08B4.5 (DNA pol E2) (c) http://www.wormbase.org/search/gene/WBGene00000423 (d) http://www.wormbase.org/search/gene/NM_060013 Direct link to protein page via UniProt ID: http://www.wormbase.org/search/protein/P91870	(Harris *et al.*, 2014)
Danio rerio	ZFIN	Home page: http://zfin.org/ Direct link to gene page via ZFIN ID: http://zfin.org/ZDB-GENE-030131-574 Query via (a) gene symbol (b) RefSeq nucleotide (c) UniProt ID: (a) http://zfin.org/action/quicksearch/query?query=tp53 (p53) (b) http://zfin.org/action/quicksearch/query?query=NM_212564 (Cdk1) (c) http://zfin.org/action/quicksearch/query?query=Q6DRD3 (DNA pol β)	(Howe *et al.*, 2013)
Dictyostelium discoideum	dictyBase	Home page: http://dictybase.org/ Direct link to Gene Summary via (a) gene name (b) gene ID (c) sequence ID (d) GenBank nucleotide (e) UniProt ID: (a) http://dictybase.org/gene/cdk1 (b) http://dictybase.org/gene/DDB_G0282191 (DNA pol A1) (c) http://dictybase.org/db/cgi-bin/search/search.pl?query=DDB0001991 (d) http://dictybase.org/db/cgi-bin/search/search.pl?query=L09720 (e) http://dictybase.org/db/cgi-bin/search/search.pl?query=P33519	(Basu *et al.*, 2013)
Drosophila melanogaster	FlyBase	Home page: http://flybase.org/ Query via gene name or symbol: http://flybase.org/cgi-bin/quicksearch_solr.cgi?data_class=FBgn&context=leonardo Direct link to Gene Report via (a) FlyBase ID (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt ID: (a) http://flybase.org/reports/FBgn0004106 (Cdk1) (b) http://flybase.org/cgi-bin/quicksearch_solr.cgi?data_class=FBgn&context=NM_079738 (DNA pol ε) (c) http://flybase.org/cgi-bin/quicksearch_solr.cgi?data_class=FBgn&context=NP_649324 (d) http://flybase.org/cgi-bin/quicksearch_solr.cgi?data_class=FBgn&context=O61661	(St Pierre *et al.*, 2014)

Supplemental Table S1 (continued)

Species-specific gene databases

Species	Resource	Example web-links	Reference
Homo sapiens	HGNC	Home page: http://www.genenames.org/ Query via (a) gene term (b) Ensembl (c) RefSeq nucleotide (d) UniProt ID: (a) http://www.genenames.org/cgi-bin/search?search=TP53 (p53) (b) http://www.genenames.org/cgi-bin/search?search=ENSG00000170312 (Cdk1) (c) http://www.genenames.org/cgi-bin/search?search=NM_006231 (DNA pol ε) (d) http://www.genenames.org/cgi-bin/search?search=Q96CS2 Direct link to Symbol Report page via HGNC ID: http://www.genenames.org/data/hgnc_data.php?hgnc_id=33830	(Gray *et al.*, 2013)
Mus musculus	MGI	Home page: http://www.informatics.jax.org/ Query via (a) gene symbol (b) RefSeq nucleotide or protein (c) UniProt ID: (a) http://www.informatics.jax.org/searchtool/Search.do?query=Cdk1 (b) http://www.informatics.jax.org/searchtool/Search.do?query=NM_011132 (DNA pol ε) (c) http://www.informatics.jax.org/searchtool/Search.do?query=Q65Z40 Direct link to Gene Detail page via MGI ID: http://www.informatics.jax.org/marker/MGI:98834 (p53)	(Drabkin and Blake, 2012)
Rattus norvegicus	RGD	Home page: http://rgd.mcw.edu/ Query via gene symbol: (a) http://rgd.mcw.edu/rgdweb/search/genes.html?term=Tp53 (p53) Direct link to Gene page via RGD ID: http://rgd.mcw.edu/rgdweb/report/gene/main.html?id=621175	(Laulederkind *et al.*, 2012)
Saccharomyces cerevisiae	SGD	Home page: http://www.yeastgenome.org/ Direct link via (a) gene name (b) systematic name: (a) http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=CDC28 (Cdk1) (b) http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=YNL262W (DNA pol 2) Query via (c) RefSeq nucleotide or protein (d) UniProt ID: (c) http://www.yeastgenome.org/cgi-bin/search/luceneQS.fpl?query=NM_001184023 (d) http://www.yeastgenome.org/cgi-bin/search/luceneQS.fpl?query=P00546	(Costanzo *et al.*, 2014)
Schizosaccharomyces pombe	PomBase	Home page: http://www.pombase.org/ Query via any term, e.g. gene name: http://www.pombase.org/search/ensembl/cdc2 (Cdk1) Direct link via (a) systematic gene ID (b) RefSeq nucleotide or protein (c) UniProt ID: (a) http://www.pombase.org/spombe/result/SPBC25H2.13c (DNA pol 2) (b) http://www.pombase.org/search/ensembl/NM_001021855 (c) http://www.pombase.org/search/ensembl/P50528	(Wood *et al.*, 2012)
Xenopus laevis, Xenopus tropicalis	Xenbase	Home page: http://www.xenbase.org/ Query via any term, e.g. gene symbol: http://www.xenbase.org/gene/searchGene.do?method=search&searchIn=0&searchType=0&searchValue=TP53 (p53) Link to Summary page via (a) GenBank (b) RefSeq nucleotide or protein (c) UniProt ID: (a) http://www.xenbase.org/gene/searchGene.do?method=search&searchIn=0&searchType=0&searchValue=BC045078 (Cdk1) (b) http://www.xenbase.org/gene/searchGene.do?method=search&searchIn=0&searchType=0&searchValue=NM_001030470 (DNA pol ε) (c) http://www.xenbase.org/gene/searchGene.do?method=search&searchIn=0&searchType=0&searchValue=P30309	(James-Zorn *et al.*, 2013)
Pathogens	GeneDB	Home page: http://www.genedb.org/ Direct link to General Information page, via gene code: http://www.genedb.org/gene/LmjF.11.0110	(Logan-Klumpler *et al.*, 2012)
Viruses (all)	NCBI Viral Genomes	Home page: http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html Direct link to record via GenBank, GI, RefSeq, UniProt ID: http://www.ncbi.nlm.nih.gov/nuccore/NC_001802.1	(Bao *et al.*, 2004)
Viruses (select human pathogens)	ViPR	Home page: http://www.viprbrc.org/ Direct link to Gene/Protein Details page, via NCBI ID: http://www.viprbrc.org/brc/viprDetails.do?ncbiAccession=AY304488	(Pickett *et al.*, 2012)

Supplemental Table S2

"Home" databases of sequence-based hits, sequence retrieval and genomic context

Resource	Type	Example web-links	Reference
NCBI Gene	Gene	Home page: http://www.ncbi.nlm.nih.gov/gene/ Gene record via NCBI Gene ID: http://www.ncbi.nlm.nih.gov/gene/7157 (human p53)	(NCBI, 2014)
NCBI Nucleotide	Nucleic acid *(all GenBank / EMBL / DDBJ entries)*	Home page: http://www.ncbi.nlm.nih.gov/nuccore/ Record via (a) RefSeq nucleotide ID (b) GI number: (a) http://www.ncbi.nlm.nih.gov/nuccore/NM_007659.3 (Cdk1) (b) http://www.ncbi.nlm.nih.gov/nuccore/442620637 (DNA pol ε) Sequence (FASTA format) via (c) RefSeq nucleotide ID (d) GI number: (c) http://www.ncbi.nlm.nih.gov/nuccore/NM_001790.4?report=fasta (d) http://www.ncbi.nlm.nih.gov/nuccore/262399398?report=fasta	(NCBI, 2014)
NCBI Protein	Protein *(includes nr-database entries)*	Home page: http://www.ncbi.nlm.nih.gov/protein/ Protein record via (a) RefSeq protein (b) GenBank (c) GI number: (a) http://www.ncbi.nlm.nih.gov/protein/NP_990595 (p53) (b) http://www.ncbi.nlm.nih.gov/protein/JAA33981 (Cdk1) (c) http://www.ncbi.nlm.nih.gov/protein/332190145 (DNA pol ε) Amino acid sequence (FASTA format) via (d) RefSeq protein (e) GenBank (f) GI-number: (d) http://www.ncbi.nlm.nih.gov/protein/NP_082106.2?report=fasta (e) http://www.ncbi.nlm.nih.gov/protein/AAD00168.1?report=fasta (f) http://www.ncbi.nlm.nih.gov/protein/147904804?report=fasta	(NCBI, 2014)
UniProt	Protein	Home page: http://www.uniprot.org/uniprot/ Protein record via UniProt ID: http://www.uniprot.org/uniprot/Q90X16 Amino acid sequence (FASTA format) via UniProt ID: http://www.uniprot.org/uniprot/P06493.fasta	(UniProt, 2014)
Ensembl	Genomic DNA, transcripts, protein	Home page: http://www.ensembl.org/ Gene record via Ensembl gene ID: http://www.ensembl.org/Gene/Summary?g=ENSLACG00000014014 (p53) Transcript record via species name plus Ensembl transcript ID: http://www.ensembl.org/Ciona_intestinalis/Transcript/Summary?t=ENSCING00000005978 (Cdk1) Protein record via species name plus Ensembl protein ID: http://www.ensembl.org/Pongo_abelii/Transcript/ProteinSummary?db=core;t=ENSPPYP00000005860 (DNA pol ε)	(Flicek *et al.*, 2014)
Gene Indices	Nucleic acid	Home page: http://compbio.dfci.harvard.edu/tgi/ Report page (includes sequence) via organism plus TC code: http://compbio.dfci.harvard.edu/cgi-bin/tgi/tc_report.pl?gudb=Frog&tc=TC417826 (p53)	(Lee *et al.*, 2005)
IPI	Protein	*IPI was discontinued in 2011, with entries integrated into UniProt. IPI cross-references in UniProt were removed in 2014; some remain in UniParc.* Query UniParc via IPI ID: http://www.uniprot.org/uniparc/?query=IPI00025087	(Kersey *et al.*, 2004)
Genomic context			
UCSC Genome Browser	Genomic loci and sequences	Home page: http://genome.ucsc.edu/ Query via species name and (a) gene symbol (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt ID: (a) http://genome.ucsc.edu/cgi-bin/hgTracks?org=X.+tropicalis&position=tp53 (b) http://genome.ucsc.edu/cgi-bin/hgTracks?hgHubConnect.destUrl=&org=Mouse&position=NM_007659.3 (Cdk1) (c) http://genome.ucsc.edu/cgi-bin/hgTracks?hgHubConnect.destUrl=&org=Rat&position=NP_001100622.2 (DNA pol ε) (d) http://genome.ucsc.edu/cgi-bin/hgTracks?hgHubConnect.destUrl=&org=Human&position=P30307 Direct link to detailed locus view via (human) chromosomal coordinates: http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr19:7151995-7294011	(Karolchik *et al.*, 2014)
Ensembl	Genomic DNA, transcripts, protein	Home page: http://www.ensembl.org/ Direct link to gene page (showing chromosomal location) via full or abbreviated species name plus (a) gene symbol (b) Ensembl gene ID (c) RefSeq nucleotide (d) RefSeq protein (e) UniProt ID: (a) http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=TP53 (p53) (b) http://www.ensembl.org/G_gorilla/Gene/Summary?g=ENSGGOG00000013311 (Cdk1) (c) http://www.ensembl.org/O_anatinus/Gene/Summary?g=XM_001514722 (DNA pol E4) (d) http://www.ensembl.org/D_rerio/Gene/Summary?g=NP_998343 (e) http://www.ensembl.org/O_cuniculus/Gene/Summary?g=G1SRF6 Link to "Region in Detail" view, via species and chromosomal coordinates: http://www.ensembl.org/mouse/Location/View?r=7:135689788-135716379	(Ensembl, 2014)
A user's guide to the human genome – article series		*Nature Genetics* (2002) – free access: http://www.nature.com/ng/journal/v32/n1s/index.html	(Wolfsberg *et al.*, 2002)

Supplemental Table S3

Ensembl Genomes databases (Kersey *et al.*, 2014)

General home page: http://ensemblgenomes.org/

Database	Example species	Example web-links
Ensembl Bacteria	*Escherichia coli*	Home page: http://bacteria.ensembl.org/ Query across all bacterial species, via any search term: http://bacteria.ensembl.org/Multi/Search/Results?species=all;q=GroEL Query (across all strains) via species name plus gene name: http://bacteria.ensembl.org/Search/Results?species=Escherichia-coli;q=dnaA (DNA pol) Query via (a) gene code (b) gene ID (c) UniProt ID: (a) http://bacteria.ensembl.org/Search/Results?q=BSU29180 (b) http://bacteria.ensembl.org/Search/Results?q=EC1011_4754 (c) http://bacteria.ensembl.org/Search/Results?q=P03004 Direct link via species name and strain plus gene ID: http://bacteria.ensembl.org/Escherichia_coli_101_1/Gene/Summary?g=EC1011_4754
Ensembl Fungi	* *Saccharomyces cerevisiae* (budding yeast), † *Schizosaccharomyces pombe* (fission yeast)	Home page: http://fungi.ensembl.org/ Query across all fungal species, via any search term: http://fungi.ensembl.org/Multi/Search/Results?species=all;q=Dam1 Direct link to gene page via species name plus (a) gene name (b) systematic name (c) RefSeq nucleotide (d) RefSeq protein (e) UniProt ID: (a*) http://fungi.ensembl.org/S_cerevisiae/Gene/Summary?g=CDC28 (Cdk1) (a†) http://fungi.ensembl.org/S_pombe/Gene/Summary?g=cdc2 (Cdk1) (b*) http://fungi.ensembl.org/S_cerevisiae/Gene/Summary?g=YNL262W (DNA pol 2) (b†) http://fungi.ensembl.org/S_pombe/Gene/Summary?g=SPBC25H2.13c (DNA pol 2) (c) http://fungi.ensembl.org/S_pombe/Gene/Summary?g=NM_001021079 (d) http://fungi.ensembl.org/S_pombe/Gene/Summary?g=NP_593305 (e) http://fungi.ensembl.org/S_pombe/Gene/Summary?g=P0CF96
Ensembl Plants	* *Arabidopsis thaliana* (thale cress), † *Oryza sativa* (rice)	Home page: http://plants.ensembl.org/ Query across all plant species, via any search term: http://plants.ensembl.org/Multi/Search/Results?species=all;q=GMI1 Direct link to gene page via species name plus (a) gene name (b) gene code (c) locus code (d) RefSeq nucleotide (e) RefSeq protein (f) UniProt ID: (a) http://plants.ensembl.org/Arabidopsis_thaliana/Gene/Summary?g=cdc2 (b*) http://plants.ensembl.org/Arabidopsis_thaliana/Gene/Summary?g=AT1G08260 (DNA pol ε) (b†) http://plants.ensembl.org/Oryza_sativa/Gene/Summary?g=P0623A10.22 (c) http://plants.ensembl.org/Oryza_sativa/Gene/Summary?g=OS06G0687600 (Cdk) (d) http://plants.ensembl.org/Arabidopsis_thaliana/Gene/Summary?g=NM_122334 (e) http://plants.ensembl.org/Arabidopsis_thaliana/Gene/Summary?g=NP_197816 (f) http://plants.ensembl.org/Oryza_sativa/Gene/Summary?g=Q7GD79
Ensembl Protists	*Dictyostelium discoideum* (slime mold), *Plasmodium falciparum* (malaria parasite), *Leishmania major* (Leishmania pathogen)	Home page: http://protists.ensembl.org/ Query across all protists, via any search term: http://protists.ensembl.org/Multi/Search/Results?species=all;q=cdc2 Direct link to gene page via species plus (a) gene symbol (b) gene code (c) RefSeq nucleotide (d) GenBank (e) UniProt ID: (a) http://protists.ensembl.org/Dictyostelium_discoideum/Gene/Summary?g=Cdk1 (b) http://protists.ensembl.org/Plasmodium_falciparum/Gene/Summary?g=PF14_0234 (DNA polymerase) (b) http://protists.ensembl.org/Leishmania_major/Gene/Summary?g=LmjF.11.0110 (c) http://protists.ensembl.org/Dictyostelium_discoideum/Gene/Summary?g=XM_630033 (d) http://protists.ensembl.org/Dictyostelium_discoideum/Gene/Summary?g=M80808 (e) http://protists.ensembl.org/Dictyostelium_discoideum/Gene/Summary?g=P33519
Ensembl Metazoa (invertebrates)	*Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm)	Home page: http://metazoa.ensembl.org/ Query across all invertebrate metazoa, via any search term: http://metazoa.ensembl.org/Multi/Search/Results?species=all;q=p53 Query via species plus descriptive name: http://metazoa.ensembl.org/Drosophila_melanogaster/Search/Results?q=string Link to gene page via species plus (a) gene name/symbol (b) gene code (c) FlyBase ID (d) WormBase ID (e) RefSeq nucleotide (f) RefSeq protein (g) GenBank protein (h) UniProt ID: (a) http://metazoa.ensembl.org/D_melanogaster/Gene/Summary?g=cdc2 (Cdk1) (a) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=cdk-1 (Cdk1) (b) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=F08B4.5 (DNA pol) (c) http://metazoa.ensembl.org/D_melanogaster/Gene/Summary?g=FBgn0004106 (d) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=WBGene00000423 (e) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=NM_060013 (f) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=NP_492414 (g) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=AAQ17186 (h) http://metazoa.ensembl.org/C_elegans/Gene/Summary?g=P91870

Supplemental Table S4

Database resources for extracting information from biomedical literature, or providing summary information about gene or protein function

Resource	Example web-links	Reference
Literature searching / information retrieval		
PubMed	Home page: http://www.ncbi.nlm.nih.gov/pubmed/ Query via any term: http://www.ncbi.nlm.nih.gov/pubmed?term=p53	(NCBI, 2014)
Google Scholar	Home page: http://scholar.google.com/ Query via any term: http://scholar.google.com/scholar?q=p53	(Boeker *et al.*, 2013)
Scirus	Query via any term: http://www.scirus.com/srsapp/search?q=p53	(Giustini and Barsky, 2005)
iHOP	Home page: http://www.ihop-net.org/UniPub/iHOP/ Query via any term (e.g. gene symbol): http://www.ihop-net.org/UniPub/iHOP/index.html?search=p53	(Fernandez *et al.*, 2007)
Textpresso ***Note:*** *need to click "Search!"*	Home page: http://www.textpresso.org/ Query via any term for (a) *C. elegans* (b) *Drosophila* (c) *Arabidopsis* (d) *Dictyostelium* (e) zebrafish (f) *S. cerevisiae* (g) *Xenopus* (h) mouse (i) rat: (a) http://www.textpresso.org/cgi-bin/celegans/search?searchstring=cdk-1 (b) http://www.textpresso.org/cgi-bin/fly/search?searchstring=stg (c) http://www.textpresso.org/cgi-bin/arabidopsis/search?searchstring=GMI1 (d) http://www.textpresso.org/cgi-bin/dicty/search?searchstring=dimB (e) http://www.textpresso.org/cgi-bin/zebrafish25/search?searchstring=plk1 (f) http://textpresso.yeastgenome.org/cgi-bin/textpresso/search?searchstring=IPL1 (g) http://www.xenbase.org/cgi-bin/textpresso/xenopus/search?searchstring=Cdc25 (h) http://www.textpresso.org/cgi-bin/mouse/search?searchstring=Apc (i) http://www.textpresso.org/cgi-bin/rat/search?searchstring=Cdk2	(Müller *et al.*, 2004)
Summary information		
NCBI Gene	Home page: http://www.ncbi.nlm.nih.gov/gene/ Query via text term e.g. gene symbol: http://www.ncbi.nlm.nih.gov/gene?term=TP53 Direct link to gene summary via (a) NCBI gene (b) Ensembl gene (c) Ensembl transcript (d) nucleotide GI number (e) Ensembl protein (f) RefSeq nucleotide or protein (g) GenBank protein (h) UniProt ID: (a) http://www.ncbi.nlm.nih.gov/gene/983#summary (b) http://www.ncbi.nlm.nih.gov/gene?term=ENSG00000156970#summary (c) http://www.ncbi.nlm.nih.gov/gene?term=ENST00000296509#summary (d) http://www.ncbi.nlm.nih.gov/gene?term=4050083[Nucleotide+UID]#summary (e) http://www.ncbi.nlm.nih.gov/gene?term=ENSP00000287598#summary (f) http://www.ncbi.nlm.nih.gov/gene?term=NM_007194#summary (g) http://www.ncbi.nlm.nih.gov/gene/?term=AAC51736#summary (h) http://www.ncbi.nlm.nih.gov/gene/?term=P06493[Nucleotide%2FProtein+Accession]#summary	(Maglott *et al.*, 2011)
UniProt	Home page: http://www.uniprot.org/ Query via any term: http://www.uniprot.org/uniprot/?query=helicase Query via species + gene symbol: http://www.uniprot.org/uniprot/?query=organism:"homo+sapiens"+gene:TP53 Query via species + gene symbol, showing only reviewed (Swiss-Prot) entries: http://www.uniprot.org/uniprot/?query=reviewed:yes+organism:"homo+sapiens"+gene:TP53 Link to General Annotation via UniProt ID: http://www.uniprot.org/uniprot/P04637#section_comments (p53)	(UniProt, 2014)
Online Mendelian Inheritance in Man (OMIM)	Home page: http://omim.org/ Query via (a) any text term (b) gene symbol: (a) http://omim.org/search?search=Bloom (b) http://omim.org/search?search=approved_gene_symbol:TP53 (p53) Query via human IDs: (c) NCBI gene (d) Ensembl gene (e) Ensembl transcript (f) NCBI nucleotide GI number (g) UniProt ID: (c) http://omim.org/search?search=gene_id:983 (Cdk1) (d) http://omim.org/search?search=ensembl_id:ENSG00000177084 (DNA pol ε) (e) http://omim.org/search?search=ensembl_id:ENST00000380502 (f) http://omim.org/search?search=ncbi_reference_sequence:345525417 (g) http://omim.org/search?search=swiss_prot_id:O96017 Query via non-human gene IDs: (h) mouse (i) zebrafish (j) *C. elegans* (k) *Drosophila*: (h) http://omim.org/search?search=mgi_id:MGI:88351 (i) http://omim.org/search?search=zfin_id:ZDB-GENE-010320-1 (j) http://omim.org/search?search=wormbase_id:WBGene00000405 (k) http://omim.org/search?search=flybase_id:FBgn0004106	(Amberger *et al.*, 2011)
Wikipedia	Home page: http://en.wikipedia.org/ Direct link to page (where available) via gene symbol: http://en.wikipedia.org/wiki/TP53	(Good *et al.*, 2012)

Supplemental Table S5

Database resources providing ontology terms to describe gene or protein function

Resource	Example web-links	Reference
Gene Ontology Project	Home page: http://www.geneontology.org/	(Blake, 2013)
Ensembl	Example GO outputs, via species name and Ensembl transcript ID: Example GO table: http://www.ensembl.org/Homo_sapiens/Transcript/Ontology/Table?t=ENST00000420246 Example GO graph: http://www.ensembl.org/Homo_sapiens/Transcript/Ontology/Image?t=ENST00000420246	(Flicek *et al.*, 2014)
QuickGO	Home page: http://www.ebi.ac.uk/QuickGO/ Query via (a) text term, e.g. gene symbol (b) Ensembl gene (c) Ensembl transcript (d) Ensembl protein ID: (a) http://www.ebi.ac.uk/QuickGO/GSearch?q=TP53 (p53) (b) http://www.ebi.ac.uk/QuickGO/GSearch?q=ENSG00000170312 (Cdk1) (c) http://www.ebi.ac.uk/QuickGO/GSearch?q=ENST00000320574 (DNA pol ε) (d) http://www.ebi.ac.uk/QuickGO/GSearch?q=ENSP00000422936 Query via species-specific gene IDs: (e) mouse (f) *Drosophila* (g) *C. elegans* (h) zebrafish (i) *Dictyostelium*: (e) http://www.ebi.ac.uk/QuickGO/GSearch?q=MGI:88351 (f) http://www.ebi.ac.uk/QuickGO/GSearch?q=FBgn0000147 (g) http://www.ebi.ac.uk/QuickGO/GSearch?q=WBGene00000405 (h) http://www.ebi.ac.uk/QuickGO/GSearch?q=ZDB-GENE-010320-1 (i) http://www.ebi.ac.uk/QuickGO/GSearch?q=DDB_G0272813 Direct link via UniProt ID: http://www.ebi.ac.uk/QuickGO/GProtein?ac=P06493	(Huntley *et al.*, 2009)
PANTHER	Home page: http://www.pantherdb.org/ Direct link to specific PANTHER record via (a) RefSeq nucleotide (b) RefSeq protein (c) UniProt ID: (a) http://www.pantherdb.org/genes/geneList.do?searchType=basic&fieldName=all&fieldValue=NM_000546 (human p53) (b) http://www.pantherdb.org/genes/geneList.do?searchType=basic&fieldName=all&fieldValue=NP_031685 (mouse Cdk1) (c) http://www.pantherdb.org/genes/geneList.do?searchType=basic&fieldName=all&fieldValue=Q19196 (*C. elegans* DNA pol ε)	(Mi *et al.*, 2013)
A list of other GO-related tools is hosted here: http://neurolex.org/wiki/Category:Resource:Gene_Ontology_Tools		

Supplemental Table S6

Database resources providing information about enzymes, pathways and systems

Resource	Example web-links	Reference
Biochemical reactions		
IntEnz	Home page: http://www.ebi.ac.uk/intenz/ Query via gene symbol: http://www.ebi.ac.uk/intenz/query?cmd=Search&q=Cdk1 Query via UniProt ID: http://www.ebi.ac.uk/intenz/query?cmd=Search&q=P06493 (human Cdk1)	(Fleischmann *et al.*, 2004)
BRENDA	Home page: http://www.brenda-enzymes.org/ Query via gene name: http://www.brenda-enzymes.org/index.php4?page=/php/search_result.php4?a=9&W[2]=Cdk1 Query via UniProt ID: http://www.brenda-enzymes.org/php/result_flat.php4?UniProtAcc=P06493 (human Cdk1)	(Schomburg *et al.*, 2013)
Enzyme Portal	Home page: http://www.ebi.ac.uk/enzymeportal/ Query via gene symbol: http://www.ebi.ac.uk/enzymeportal/search?searchparams.type=KEYWORD&searchparams.previoustext=&searchparams.start=0&searchparams.text=PKM Direct link to enzyme page via UniProt ID: http://www.ebi.ac.uk/enzymeportal/search/P06493/enzyme (human Cdk1)	(Alcántara *et al.*, 2013)
Signaling pathways		
BioCarta Pathways	Home page: http://www.biocarta.com/genes/index.asp *Example pathway diagram "Cell Cycle: G1/S Check Point":* http://www.biocarta.com/pathfiles/h_g1Pathway.asp	(Nishimura, 2001)
KEGG PATHWAY	Home page: http://www.genome.jp/kegg/pathway.html Query pathways via gene symbol: http://www.genome.jp/kegg-bin/search_pathway_text?map=map&mode=1&viewImage=true&keyword=TP53 (p53) *Example pathway diagram "CELL CYCLE", with p53 highlighted:* http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map04110&keyword=TP53	(Kanehisa *et al.*, 2014)
Reactome	Home page: http://www.reactome.org/ Query via (a) gene symbol (b) Ensembl gene, RefSeq nucleotide or protein, or UniProt ID: (a) http://www.reactome.org/cgi-bin/search2?DB=gk_current&OPERATOR=ALL&QUERY=TP53 (p53) (b) http://www.reactome.org/cgi-bin/search2?DB=gk_current&OPERATOR=ALL&QUERY=NM_006231 (DNA pol ε) *Example pathway diagram* "Cell Cycle Checkpoints*"; highlighted section,* "p53-Dependent G1 DNA Damage Response": http://www.reactome.org/entitylevelview/PathwayBrowser.html#DB=gk_current&FOCUS_SPECIES_ID=48887&FOCUS_PATHWAY_ID=69620&ID=69563	(Croft *et al.*, 2014)
WikiPathways	Home page: http://wikipathways.org/ Query via (a) gene name/symbol (b) organism name plus gene symbol: (a) http://wikipathways.org//index.php?title=Special%3ASearchPathways&doSearch=1&query=CDK1 (b) http://wikipathways.org//index.php?title=Special%3ASearchPathways&doSearch=1&species=Homo+sapiens&query=TP53	(Kelder *et al.*, 2012)
Pathway Commons	Home page: http://www.pathwaycommons.org/ Query pathways via gene symbol or UniProt ID: http://www.pathwaycommons.org/pc/webservice.do?version=3.0&format=html&cmd=get_by_keyword&snapshot_id=GLOBAL_FILTER_SETTINGS&record_type=PATHWAY&q=P04637 Link to Network Visualizer via gene symbol or UniProt ID: http://www.pathwaycommons.org/pcviz/#neighborhood/ABL1	(Cerami *et al.*, 2011)
Biological Systems		
BioSystems (NCBI)	Home page: http://www.ncbi.nlm.nih.gov/biosystems/ Query via (a) gene symbol (b) RefSeq or GenBank protein (c) UniProt ID: (a) http://www.ncbi.nlm.nih.gov/biosystems/?term=TP53 (p53) (b) http://www.ncbi.nlm.nih.gov/biosystems/?term=NP_031685 (Cdk1) (c) http://www.ncbi.nlm.nih.gov/biosystems/?term=Q9H410	(Geer *et al.*, 2010)

Supplemental Table S7

Database resources providing primary information on gene or protein function from specific research projects

Resource	Example web-links	Reference
Specific research projects		
S. pombe deletion screen *(PomBase)*	Home page: http://www.pombase.org/ Query via any term, e.g. gene name: http://www.pombase.org/search/ensembl/cdc2 (Cdk1) Direct link to gene page via (a) systematic gene ID (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt ID: (a) http://www.pombase.org/spombe/result/SPBC25H2.13c (DNA pol 2) (b) http://www.pombase.org/search/ensembl/NM_001021855 (c) http://www.pombase.org/search/ensembl/NP_593305 (d) http://www.pombase.org/search/ensembl/P30290	(Kim *et al.*, 2010; Wood *et al.*, 2012; Hayles *et al.*, 2013)
C. elegans RNAi screen *(PhenoBank)*	Home page: http://www.worm.mpi-cbg.de/phenobank/ Direct link to Gene Summary page via *C. elegans* sequence code: http://www.worm.mpi-cbg.de/phenobank/cgi-bin/GenePage.py?Gene=T05G5.3 (Cdk1)	(Sönnichsen *et al.*, 2005)
Human RNAi screens, mitotic protein interaction & localization data *(MitoCheck)*	Home page: http://www.mitocheck.org/cgi-bin/mtc/ Query via (a) generic gene or protein term (b) gene symbol: (a) http://www.mitocheck.org/cgi-bin/mtc?query=importin (b) http://www.mitocheck.org/cgi-bin/mtc?query=TP53 (p53) Direct link to gene page via human (c) Ensembl gene (d) UniProt ID: (c) http://www.mitocheck.org/cgi-bin/mtc?query=ENSG00000170312 (Cdk1) (d) http://www.mitocheck.org/cgi-bin/mtc?query=Q07864 (DNA pol ε) Query via non-human orthologue (e) gene name (f) Ensembl gene ID: (e) http://www.mitocheck.org/cgi-bin/mtc?query=IPL1 (f) http://www.mitocheck.org/cgi-bin/mtc?query=ENSDARG00000087554	(Hutchins *et al.*, 2010; Neumann *et al.*, 2010)

Supplemental Table S8

Resources for identifying homologues and determining evolutionary conservation

Resource	Example web-links	Reference
Listings of pre-identified homologues		
Ensembl Orthologues	Link via Ensembl gene ID: http://www.ensembl.org/Gene/Compara_Ortholog?g=ENSG00000158402 Link via species abbreviation plus (a) gene symbol (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt ID: (a) http://www.ensembl.org/H_sapiens/Gene/Compara_Ortholog?g=CDK1 (b) http://www.ensembl.org/M_musculus/Gene/Compara_Ortholog?g=NM_027830 (c) http://www.ensembl.org/X_tropicalis/Gene/Compara_Ortholog?g=NP_001107968 (d) http://www.ensembl.org/D_rerio/Gene/Compara_Ortholog?g=F1QDI9	(Flicek *et al.*, 2014)
Ensembl Paralogues	Link via Ensembl gene ID: http://www.ensembl.org/H_sapiens/Gene/Compara_Paralog?g=ENSG00000164045 Link via species abbreviation plus (a) gene symbol (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt ID: (a) http://www.ensembl.org/H_sapiens/Gene/Compara_Paralog?g=AURKA (b) http://www.ensembl.org/M_musculus/Gene/Compara_Paralog?g=NM_015733 (c) http://www.ensembl.org/R_norvegicus/Gene/Compara_Paralog?g=NP_598256 (d) http://www.ensembl.org/D_rerio/Gene/Compara_Paralog?g=P79734	(Flicek *et al.*, 2014)
NCBI HomoloGene	Home page: http://www.ncbi.nlm.nih.gov/homologene/ Direct link via (a) gene symbol (b) NCBI gene ID (c) RefSeq nucleotide (d) RefSeq protein: (a) http://www.ncbi.nlm.nih.gov/homologene?term=TP53 (p53) (b) http://www.ncbi.nlm.nih.gov/homologene?term=983[GeneID] (human Cdk1) (c) http://www.ncbi.nlm.nih.gov/homologene?term=XM_001233984 (chicken DNA pol ε) (d) http://www.ncbi.nlm.nih.gov/homologene?term=NP_035145	(NCBI, 2014)
Retrieval of similar-sequence protein entries		
UniProt Reference Clusters (UniRef)	UniRef100 (100% identity), via UniProt ID: http://www.uniprot.org/uniprot/?query=cluster:(member:P04637+identity:1.0) (human p53) UniRef90 (90% identity), via RefSeq mRNA ID: http://www.uniprot.org/uniprot/?query=cluster:(uniprot:(NM_007659)+identity:0.9) (mouse Cdk1) UniRef50 (50% identity), via RefSeq protein ID: http://www.uniprot.org/uniprot/?query=cluster:(uniprot:(NP_001100622)+identity:0.5) (rat DNA pol ε)	(Suzek *et al.*, 2007; UniProt, 2014)
Retrieval of phylogenetic trees		
TreeFam	Home page: http://www.treefam.org/ Direct link to Summary view via: (a) gene symbol (b) Ensembl gene (c) FlyBase gene (d) Ensembl protein (e) UniProt ID: (a) http://www.treefam.org/family/TP53 (p53) (b) http://www.treefam.org/family/ENSMUSG00000019942 (mouse Cdk1) (c) http://www.treefam.org/family/FBgn0264326 (*Drosophila* DNA pol ε) (d) http://www.treefam.org/family/ENSP00000428982 (e) http://www.treefam.org/family/P06493 Direct link to Gene Tree view via: (f) gene symbol (g) Ensembl gene (h) FlyBase gene (i) Ensembl protein (j) UniProt ID: (f) http://www.treefam.org/family/BRCA2#tabview=tab1 (g) http://www.treefam.org/family/ENSMUSG00000032113#tabview=tab1 (h) http://www.treefam.org/family/FBgn0003124#tabview=tab1 (i) http://www.treefam.org/family/ENSP00000428982#tabview=tab1 (j) http://www.treefam.org/family/P06493#tabview=tab1	(Schreiber *et al.*, 2014)
Launch of sequence similarity searches		
BLAST at NCBI ***Note:*** *need to click "BLAST".*	Home page: http://blast.ncbi.nlm.nih.gov/ Nucleotide via RefSeq mRNA ID: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&QUERY=NM_001071775.2 Protein via RefSeq protein ID: http://blast.ncbi.nlm.nih.gov/Blast.cgi?BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&QUERY=NP_524860	(NCBI, 2014)
BLAST protein at UniProt	Home page: http://www.uniprot.org/?tab=blast Launch via UniProt ID: http://www.uniprot.org/blast/?query=P04637 (p53)	(UniProt, 2014)

Supplemental Table S9

Database resources providing information about gene expression

Resource	Example web-links	Reference
SPELL (*S. cerevisiae*)	Home page: http://spell.princeton.edu/spell/ Query via (a) standard gene name (b) systematic gene name: (a) http://spell.princeton.edu/spell/search/show_results?search_string=CDC28 (Cdk1) (b) http://spell.princeton.edu/spell/search/show_results?search_string=YPL209C	(Hibbs *et al.*, 2007)
Gene Expression Omnibus	Home page: http://www.ncbi.nlm.nih.gov/geoprofiles/ Query via (a) text term, e.g. gene symbol (b) RefSeq nucleotide ID (c) NCBI nucleotide GI number: (a) http://www.ncbi.nlm.nih.gov/geoprofiles?term=TP53 (p53) (b) http://www.ncbi.nlm.nih.gov/geoprofiles?term=NM_001071775 (c) http://www.ncbi.nlm.nih.gov/geoprofiles/?term=38014291[GI]	(Barrett *et al.*, 2013)
ArrayExpress	Home page: http://www.ebi.ac.uk/arrayexpress/ Query via (a) text term (b) organism name plus gene symbol: (a) http://www.ebi.ac.uk/arrayexpress/browse.html?keywords=cyclin (b) http://www.ebi.ac.uk/arrayexpress/browse.html?keywords=Homo+sapiens+TP53	(Rustici *et al.*, 2013)
Gene Expression Atlas	Home page: http://www.ebi.ac.uk/gxa/ Query via text term, e.g. gene symbol (across species): http://www.ebi.ac.uk/gxa/query?geneQuery=RAD21 Direct link via Ensembl gene ID: http://www.ebi.ac.uk/gxa/genes/ENSG00000141510 (human p53) Direct link via mouse (MGI) gene ID: http://www.ebi.ac.uk/gxa/query?geneQuery=MGI:88351 (Cdk1) Direct link via *C. elegans* gene code: http://www.ebi.ac.uk/gxa/genes/K08E3.6 (DNA pol ε) Direct link via FlyBase ID: http://www.ebi.ac.uk/gxa/genes/FBgn0000147 Direct link via Ensembl transcript ID: http://www.ebi.ac.uk/gxa/genes/ENST00000509894 Direct link via RefSeq nucleotide ID: http://www.ebi.ac.uk/gxa/query?geneQuery=NM_000546 Direct link via RefSeq protein ID: http://www.ebi.ac.uk/gxa/query?geneQuery=NP_009125 Direct link via UniProt ID: http://www.ebi.ac.uk/gxa/query?geneQuery=P04637	(Kapushesky *et al.*, 2012)
The Human Protein Atlas	Home page: http://www.proteinatlas.org/ Query via text term (e.g. gene symbol): http://www.proteinatlas.org/search/TP53 (p53) Query via human Ensembl transcript ID: http://www.proteinatlas.org/search/ENST00000395284 (Cdk1) Query via human Ensembl protein ID: http://www.proteinatlas.org/search/ENSP00000322570 (DNA pol ε) Query via human UniProt ID: http://www.proteinatlas.org/search/O60566 Direct link to database entry via human Ensembl gene ID: http://www.proteinatlas.org/ENSG00000156970	(Asplund *et al.*, 2012)

Supplemental Table S10

Database resources: protein domains and short linear motifs (SLiMs)

Resource	Example web-links	Reference
Annotation of protein domains		
Pfam	Home page: http://pfam.sanger.ac.uk/ Query via UniProt name: http://pfam.sanger.ac.uk/protein/P53_HUMAN Query via UniProt ID: http://pfam.sanger.ac.uk/protein/G3SP81	(Finn *et al.*, 2014)
InterPro	Home page: http://www.ebi.ac.uk/interpro/ Query via any term: http://www.ebi.ac.uk/interpro/search?q=Cdk1 Link via UniProt ID: http://www.ebi.ac.uk/interpro/protein/P04637 (p53)	(Hunter *et al.*, 2012)
SMART	Home page: http://smart.embl.de/ Direct link via (a) Ensembl protein ID (b) UniProt name (c) UniProt ID (d) UniProt ID, *including Pfam domains:* (a) http://smart.embl.de/smart/show_motifs.pl?ID=ENSP00000322570 (DNA pol ε) (b) http://smart.embl.de/smart/show_motifs.pl?ID=CDK1_MOUSE (c) http://smart.embl.de/smart/show_motifs.pl?ID=P04637 (human p53) (d) http://smart.embl.de/smart/show_motifs.pl?DO_PFAM=DO_PFAM&ID=P04637	(Letunic *et al.*, 2012)
Gene3D	Home page: http://gene3d.biochem.ucl.ac.uk/ Query via (a) gene symbol (b) Ensembl gene (c) RefSeq nucleotide or protein (d) UniProt ID: (a) http://gene3d.biochem.ucl.ac.uk/search?mode=protein&sterm=CHEK1 (b) http://gene3d.biochem.ucl.ac.uk/search?mode=protein&sterm=ENSG00000156970 (c) http://gene3d.biochem.ucl.ac.uk/search?mode=protein&sterm=NM_001790 (d) http://gene3d.biochem.ucl.ac.uk/search?mode=protein&sterm=P04637	(Lees *et al.*, 2014)
Prediction of SLiMs		
PROSITE	Home page: http://prosite.expasy.org/ Direct link via UniProt ID: http://www.expasy.org/cgi-bin/prosite/PSScan.cgi?seq=P04637 (p53) Direct link via UniProt name: http://www.expasy.org/cgi-bin/prosite/PSScan.cgi?seq=RCC1_HUMAN	(Sigrist *et al.*, 2013)
Minimotif Miner	Home page: http://minimotifminer.org	(Mi *et al.*, 2012)
Scansite	Home page: http://scansite.mit.edu/ Direct link via UniProt ID: http://scansite.mit.edu/cgi-bin/motifscan_id?database=swissprot&motif_option=all&stringency=High&protein_id=P04637 (p53)	(Obenauer *et al.*, 2003)
Annotation of protein domains and SLiMs		
Conserved Domain Database (CDD)	Home page: http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml Direct link via (a) RefSeq mRNA (b) RefSeq protein (c) UniProt name (d) UniProt ID: (a) http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=NP_000537.3 (p53) (b) http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=NP_035145 (c) http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=RAD53_YEAST (d) http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=Q8R4E9	(Marchler-Bauer *et al.*, 2013)
Eukaryotic Linear Motif (ELM) Functional Site Prediction	Home page: http://elm.eu.org/ Direct link via (a) UniProt ID (b) UniProt name: (a) http://elm.eu.org/cgimodel.py?fun=Submit&swissprotId=P04637 (p53) (b) http://elm.eu.org/cgimodel.py?fun=Submit&swissprotId=BUB1B_HUMAN	(Dinkel *et al.*, 2014)
DASty	Home page: http://www.ebi.ac.uk/dasty/ Direct link via UniProt ID: http://www.ebi.ac.uk/dasty/client/index.html?q=P04637 (p53) Direct link via UniProt protein name: http://www.ebi.ac.uk/dasty/client/index.html?q=BRCA1_HUMAN	(Villaveces *et al.*, 2011)
ANNIE > Click "Interactive View"	Home page: http://annie.bii.a-star.edu.sg/ Direct link via UniProt ID: http://annie.bii.a-star.edu.sg/annie/query.do?seqid=P04637 (p53)	(Ooi *et al.*, 2009)

Supplemental Table S11

Database resources providing information about protein 3D structure

The example PDB entry "1tup" corresponds to the structure "Tumor suppressor p53 complexed with DNA".

Resource	Example web-links	Reference
Summary information about protein structures		
RCSB PDB	Home page: http://www.pdb.org/ Query via UniProt ID: http://www.pdb.org/pdb/search/smart.do?smartSearchSubtype_0=UpAccessionIdQuery&accessionIdList_0=P04637 (p53) Direct link to "Protein Feature View" via UniProt ID: http://www.rcsb.org/pdb/protein/P04637 (p53) Direct link to Structure Summary page via PDB ID: http://www.rcsb.org/pdb/explore/explore.do?structureId=1tup	(Rose *et al.*, 2013)
PDB in Europe (PDBe)	Home page: http://www.ebi.ac.uk/pdbe/ Query via text term (e.g. gene symbol): http://www.ebi.ac.uk/pdbe/searchResults.html?term=TP53 (p53) Query via any UniProt ID: http://www.ebi.ac.uk/pdbe/searchResults.html?term=P24941 (Cdk2) Direct link to PDBeView Summary page via PDB ID: http://www.ebi.ac.uk/pdbe-srv/view/entry/1tup/summary.html	(Gutmanas *et al.*, 2014)
NCBI Molecular Modeling Database (MMDB)	Home page: http://www.ncbi.nlm.nih.gov/structure/ Query via text term (e.g. gene symbol): http://www.ncbi.nlm.nih.gov/structure/?term=TP53 (p53) Direct link to Protein Structure Summary page via PDB ID: http://www.ncbi.nlm.nih.gov/structure/?term=1tup	(Madej *et al.*, 2012)
PDBsum	Home page: http://www.ebi.ac.uk/pdbsum/ Direct link to alignment via UniProt ID: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetUnichains.pl?uniprot_id=P04637 (p53) Direct link to alignment via UniProt protein name: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetUnichains.pl?uniprot_code=CHK1_HUMAN Direct link to "Top page" protein structure and function summary, via PDB ID: http://www.ebi.ac.uk/pdbsum/1tup	(de Beer *et al.*, 2014)
UniPDB	Home page: http://www.ebi.ac.uk/pdbe-apps/widgets/unipdb/ Direct link to alignment via UniProt ID: http://www.ebi.ac.uk/pdbe-apps/widgets/unipdb?uniprot=P04637 (p53) Direct link to alignment via UniProt name: http://www.ebi.ac.uk/pdbe/widgets/unipdb?uniprot=CHK1_HUMAN	(Gutmanas *et al.*, 2014)
Web-launchable interactive 3D structure viewers		
Jmol Viewer	Home page: http://jmol.sourceforge.net/ User guide: http://jmol.sourceforge.net/#Learn%20to%20use%20Jmol Direct link to 3D structure in viewer (from RCSB), via PDB ID: http://www.rcsb.org/pdb/explore/jmol.do?structureId=1tup	(Herraez, 2006)
Cn3D Viewer	Home page: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml Software download and install: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dinstall.shtml Tutorial: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dtut.shtml Direct link to 3D structure in viewer (from MMDB), via PDB ID: http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?db=t&form=6&uid=1tup&Dopt=i *(requires the Cn3D program to be pre-installed)*	(Wang *et al.*, 2000)
AstexViewer™	Home page: http://astx.com/technology/pyramid-platform/#astexviewer Documentation (at PDBe): http://www.ebi.ac.uk/pdbe-srv/view/viewer/ViewerDocumentation.html Direct link to 3D structure in viewer (from PDBe), via PDB ID: http://www.ebi.ac.uk/pdbe-srv/view/entry/1tup/viewer	(Hartshorn, 2002)
OpenAstexViewer	Home page: http://openastexviewer.net/web/ User interface guide: http://openastexviewer.net/web/interface.html Direct link to 3D structure in viewer (from PDBe), via PDB ID: http://www.ebi.ac.uk/pdbe-srv/view/entry/1tup/openastex	(Oldfield, 2004)

Supplemental Table S12

Database resources providing information about protein interactors

Resource	Example web-links	Reference
STRING	Home page: http://string-db.org/ Query via text term: http://string-db.org/newstring_cgi/show_network_section.pl?identifier=p53 Direct link to interaction diagram via taxonomy code plus gene symbol: http://string-db.org/newstring_cgi/show_network_section.pl?species=9606&identifier=CDK1 Direct link to interaction diagram via (a) Ensembl gene ID (b) UniProt ID: (a) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=ENSG00000156970 (b) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=P14635 Direct link to interaction diagram via gene IDs for species databases: (c) FlyBase (d) WormBase (e) SGD (f) PomBase: (c) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=FBgn0000147 (d) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=K08E3.6 (e) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=YBR156C (f) http://string-db.org/newstring_cgi/show_network_section.pl?identifier=SPAC24H6.05	(Franceschini *et al.*, 2013)
IntAct	Home page: http://www.ebi.ac.uk/intact Direct link via (a) gene symbol (b) Ensembl gene ID (c) UniProt ID: (a) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=TP53 (b) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=P14635 (c) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=ENSG00000156970 Direct link via gene IDs for species databases: (d) FlyBase (e) WormBase (f) SGD (g) PomBase (h) TAIR: (d) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=FBgn0000147 (e) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=K08E3.6 (f) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=YBR156C (g) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=SPAC24H6.05 (h) http://www.ebi.ac.uk/intact/pages/interactions/interactions.xhtml?query=AT1G15570	(Kerrien *et al.*, 2012)
BioGRID	Home page: http://thebiogrid.org/ Query via text (e.g. gene symbol): http://thebiogrid.org/search.php?search=TP53 (p53) Direct link via sequence IDs: (a) Ensembl gene ID (b) RefSeq mRNA (c) RefSeq protein (d) UniProt: (a) http://thebiogrid.org/search.php?search=ENSMUSG00000019942 (Cdk1) (b) http://thebiogrid.org/search.php?search=NM_001790 (c) http://thebiogrid.org/search.php?search=NP_035145 (d) http://thebiogrid.org/search.php?search=Q96GD4 Direct link via gene IDs for species databases: (e) NCBI gene (f) FlyBase (g) WormBase (h) SGD (i) PomBase (j) TAIR (k) dictyBase (l) ZFIN: (e) http://thebiogrid.org/search.php?search=1111 (f) http://thebiogrid.org/search.php?search=FBgn0000147 (g) http://thebiogrid.org/search.php?search=K08E3.6 (h) http://thebiogrid.org/search.php?search=YBR156C (i) http://thebiogrid.org/search.php?search=SPAC24H6.05 (j) http://thebiogrid.org/search.php?search=AT5G24280 (k) http://thebiogrid.org/search.php?search=DDB_G0272813 (l) http://thebiogrid.org/search.php?search=ZDB-GENE-010320-1	(Chatr-Aryamontri *et al.*, 2013)
IMEx Consortium	Home page: http://www.imexconsortium.org/ Cross-database query via (a) gene symbol (b) UniProt ID: (a) http://www.ebi.ac.uk/intact/imex/main.xhtml?query=TP53 (b) http://www.ebi.ac.uk/intact/imex/main.xhtml?query=P11440 (Cdk1)	(Orchard *et al.*, 2012)

Supplemental Table S13

Database resources providing information about protein post-translational modifications

Resource	Example web-links	Reference
PhosphoSitePlus	Home page: http://www.phosphosite.org/ Query via text term: http://www.phosphosite.org/simpleSearchSubmitAction.do?queryId=-1&from=0&searchStr=sororin Direct link via UniProt ID: http://www.phosphosite.org/uniprotAccAction.do?id=P04637 (p53)	(Hornbeck *et al.*, 2012)
Phospho.ELM	Home page: http://phospho.elm.eu.org/ Direct link via Ensembl protein ID: http://phospho.elm.eu.org/byAccession/ENSP00000269305 (p53) Direct link via UniProt ID: http://phospho.elm.eu.org/byAccession/P06493 (Cdk1) Direct link via UniProt name: http://phospho.elm.eu.org/byAccession/CASP9_HUMAN	(Dinkel *et al.*, 2011)
PHOSIDA	Home page: http://www.phosida.com/ Query via UniProt ID: http://141.61.102.18/phosida/index.aspx?query=P04637 (p53)	(Gnad *et al.*, 2011)
UbiProt	Home page: http://ubiprot.org.ru/ Direct link via UniProt ID: http://ubiprot.org.ru/index.php?mode=proteins_show&spac=Q9Y6K9	(Chernorudskiy *et al.*, 2007)
dbPTM	Home page: http://dbptm.mbc.nctu.edu.tw/ Direct link via UniProt protein name: http://dbptm.mbc.nctu.edu.tw/search_result.php?search_type=db_id&swiss_id=P53_HUMAN (p53)	(Lu *et al.*, 2013)

Supplemental Table S14

Database resources: variations and mutations in genes and proteins

Resource	Example web-links	Reference
dbSNP and dbVar	Home page (dbSNP): http://www.ncbi.nlm.nih.gov/SNP/ Home page (dbVar): http://www.ncbi.nlm.nih.gov/dbvar/ Direct link to dbVar table via gene symbol: http://www.ncbi.nlm.nih.gov/dbvar/?term=TP53[Gene] (human p53) Direct link to Variation Viewer via gene symbol: http://www.ncbi.nlm.nih.gov/sites/varvu?gene=TP53 (human p53) Direct link to Variation Viewer via NCBI GeneID: http://www.ncbi.nlm.nih.gov/sites/varvu?gene=983 (human Cdk1)	(NCBI, 2014)
Ensembl	Home page: http://www.ensembl.org/ "About Ensembl Variation" documentation page: http://www.ensembl.org/info/docs/variation/index.html Direct link to Variation Table, via species name and gene symbol: http://www.ensembl.org/H_sapiens/Gene/Variation_Gene/Table?g=CFTR Direct link to Variation Table, via species name and UniProt ID: http://www.ensembl.org/M_musculus/Gene/Variation_Gene/Table?g=P11440 Direct link to Variation Table, via Ensembl gene ID: http://www.ensembl.org/Gene/Variation_Gene/Table?g=ENSG00000141510 (human p53) Direct link to Variation Table, via Ensembl transcript ID: http://www.ensembl.org/Gene/Variation_Gene/Table?g=ENSMUST00000020099 (mouse Cdk1) Direct link to Variation Table, via Ensembl protein ID: http://www.ensembl.org/Gene/Variation_Gene/Table?g=ENSGALP00000012408 (chicken DNA pol ε)	(Flicek *et al.*, 2014)
UniProt	Home page: http://www.uniprot.org/ UniProt manual "Natural Variant" page: http://www.uniprot.org/manual/variant Search for reviewed entries via species plus gene name: http://www.uniprot.org/uniprot/?query=reviewed%3Ayes+Mus+musculus+gene:tp53 Search for reviewed entries via species plus any ID: http://www.uniprot.org/uniprot/?query=reviewed%3Ayes+Mus+musculus+NM_001127259 Direct link to "Sequence annotation (Features)" section, via UniProt ID: http://www.uniprot.org/uniprot/P04637#section_features (human p53)	(UniProt, 2014)

Supplemental Table S15

Database resources: genetic association with human disease

Resource	Example web-links	Reference
UniProt	Home page: http://www.uniprot.org/ Search for reviewed human entries via any term (text or ID): http://www.uniprot.org/uniprot/?query=reviewed:yes+9606+Bcr-Abl Link to "General annotation (Comments)" section via UniProt ID (see "Involvement in disease" subsection): http://www.uniprot.org/uniprot/P04637#section_comments (p53)	(UniProt, 2014)
Online Mendelian Inheritance in Man (OMIM)	Home page: http://omim.org/ Query via (a) any text term (b) gene symbol: (a) http://omim.org/search?search=Bloom (b) http://omim.org/search?search=approved_gene_symbol:MZT1 Query via human IDs: (c) NCBI gene (d) Ensembl gene or transcript (f) NCBI nucleotide GI number (g) UniProt: (c) http://omim.org/search?search=gene_id:7157 (p53) (d) http://omim.org/search?search=ensembl_id:ENSG00000183765 (e) http://omim.org/search?search=ncbi_reference_sequence:345525417 (f) http://omim.org/search?search=swiss_prot_id:O96017 Query via gene IDs for non-human orthologues, from: (g) mouse (h) zebrafish (i) *C. elegans* (j) *Drosophila*: (g) http://omim.org/search?search=mgi_id:MGI:88351 (Cdk1) (h) http://omim.org/search?search=zfin_id:ZDB-GENE-010320-1 (i) http://omim.org/search?search=wormbase_id:WBGene00000405 (j) http://omim.org/search?search=flybase_id:FBgn0004106	(Amberger *et al.*, 2011)
KEGG	*KEGG GENES* Home page: http://www.genome.jp/kegg/genes.html Direct link via (a) NCBI gene ID (b) gene symbol: (a) http://www.genome.jp/dbget-bin/www_bget?hsa:7157 (p53) (b) http://www.genome.jp/dbget-bin/www_bget?hsa:APC *KEGG DISEASE* Home page: http://www.genome.jp/kegg/disease/ Query via text term (e.g. gene symbol): http://www.kegg.jp/medicus-bin/search?display=disease&q=CFTR	(Kanehisa *et al.*, 2014)
Genetic Association Database (GAD)	Home page: http://geneticassociationdb.nih.gov/ Direct link to disease-association table via gene symbol: http://geneticassociationdb.nih.gov/cgi-bin/tableview.cgi?associa=Y&cond=gene='TP53' (p53) Direct link to disease-association table via NCBI Gene ID: http://geneticassociationdb.nih.gov/cgi-bin/tableview.cgi?associa=Y&cond=LOCUSNUM=675	(Becker *et al.*, 2004)
Comparative Toxicogenomics Database (CTD)	Home page: http://ctdbase.org/ Direct link to gene record via gene symbol (need to click on "Diseases" tab): http://ctdbase.org/basicQuery.go?bqCat=gene&bq=name:TP53 (p53) Direct link to disease-association table via NCBI gene ID: http://ctdbase.org/detail.go?type=gene&view=disease&acc=7157 (p53)	(Davis *et al.*, 2013)
ClinVar	Home page: http://www.ncbi.nlm.nih.gov/clinvar/ Direct link to variation table via (a) gene symbol (b) NCBI gene ID: (a) http://www.ncbi.nlm.nih.gov/clinvar/?term=TP53[gene] (p53) (b) http://www.ncbi.nlm.nih.gov/clinvar?term=11200[GeneID]	(Landrum *et al.*, 2014)
DECIPHER *(development)*	Home page: https://decipher.sanger.ac.uk/ Query via text term: https://decipher.sanger.ac.uk/search?q=FANCD2	(Bragin *et al.*, 2014)
COSMIC *(cancer)*	Home page: http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/ Direct link to overview page via human gene symbol: http://cancer.sanger.ac.uk/cosmic/gene/overview?ln=APC Query via human (a) Ensembl gene (b) UniProt ID: (a) http://cancer.sanger.ac.uk/cosmic/gene/overview?ln=ENSG00000183765 (b) http://cancer.sanger.ac.uk/cosmic/search?q=P04637 (p53)	(Forbes *et al.*, 2011)
dbDEPC *(cancer)*	Home page: http://lifecenter.sgst.cn/dbdepc/index.do Direct link via UniProt ID: http://lifecenter.sgst.cn/dbdepc/detailProtein.do?uniprotKB=P04637	(He *et al.*, 2012)

Supplemental Table S16

Database resources providing information on drugs and inhibitors

Resource	Example web-links	Reference
Databases searchable by target IDs		
ChEMBL	Home page: https://www.ebi.ac.uk/chembl/ *Options for ChEMBL target-based searching include:* (a) via the ChEMBL home page: paste the hit ID into the search box, click [Targets], then the ChEMBL ID corresponding to the relevant gene product. (b) within a UniProt entry: go to the Cross-references / Chemistry section, then click the ChEMBL link. In the Target Report Card, under Target Associated Bioactivities click the desired property in the pie-chart to show a table of relevant compounds.	(Bento *et al.*, 2014)
DrugBank	Home page: http://drugbank.ca/ Search for drug targets via (a) gene name (b) UniProt name (c) UniProt ID: (a) http://www.drugbank.ca/search?utf8=%E2%9C%93&search_type=targets&query=ABL1 (b) http://www.drugbank.ca/search?utf8=%E2%9C%93&search_type=targets&query=CHK1_HUMAN (c) http://www.drugbank.ca/search?utf8=%E2%9C%93&search_type=targets&query=A9UF02 At the Search Results page, click the button corresponding to the relevant gene product (starting BE…).	(Law *et al.*, 2014)
Comparative Toxicogenomics Database (CTD)	Home page: http://ctdbase.org/ Direct link to gene record, showing the ten "Top Interacting Chemicals", via (a) gene symbol, (b) NCBI Gene ID: (a) http://ctdbase.org/basicQuery.go?bqCat=gene&bq=name:ABL1 (b) http://ctdbase.org/detail.go?type=gene&acc=983 (Cdk1) Direct link to "Chemical Interactions" table (sorted by decreasing number of references) via NCBI Gene ID: http://ctdbase.org/detail.go?sort=refSort&view=ixn&type=gene&acc=983 (Cdk1)	(Davis *et al.*, 2013)
canSAR	Home page: https://cansar.icr.ac.uk/ Direct link to "Molecular Target Synopsis – Chemistry" page, via UniProt ID: https://cansar.icr.ac.uk/cansar/molecular-targets/target_report_chemistry/P04637/	(Bulusu *et al.*, 2014)
PubChem BioAssay	Home page: http://www.ncbi.nlm.nih.gov/pcassay/ Query via (a) gene symbol (b) UniProt ID (c) protein GI number: (a) http://www.ncbi.nlm.nih.gov/pcassay?term=ABL[GeneSymbol] (b) http://www.ncbi.nlm.nih.gov/pcassay?term=Q96GD4 (c) http://www.ncbi.nlm.nih.gov/pcassay/?term=85681908[GI]	(Wang *et al.*, 2014)
Other databases of compounds and pharmaceuticals		
PubChem Compound	Home page: http://www.ncbi.nlm.nih.gov/pccompound/ Query via text term *(note: search results may not necessarily be drugs targeting the queried term)*: http://www.ncbi.nlm.nih.gov/pccompound?term=CDK1	(Wang *et al.*, 2009)
Harvester Chemical	Home page: http://harvester.kit.edu/jss/special/chemical?searchTerm= Query via any text term (e.g. gene name): http://harvester.kit.edu/jss/special/chemical?searchTerm=CDK1	
Druggability prediction		
DrugEBIlity	Home page: https://www.ebi.ac.uk/chembl/drugebility/ Direct link to protein page with druggability scores, via UniProt ID: https://www.ebi.ac.uk/chembl/drugebility/protein/P00519	
canSAR	Home page: https://cansar.icr.ac.uk/ Direct link to "Molecular Target Synopsis – Druggability" page, via UniProt ID: https://cansar.icr.ac.uk/cansar/molecular-targets/target_report_druggability/P04637/	(Bulusu *et al.*, 2014)
Prediction of docking interactions between proteins and small molecules		
Drugable	Home page: http://drugable.com/ Query via text term (e.g. gene symbol): http://drugable.com/search?query=ABL1 Direct link via UniProt ID: http://drugable.com/search?query=P00519	(Reardon, 2013)

Supplemental Table S17

Resources providing cross-database searches

Resource	Example web-links	Reference
GQuery Cross-database search	Home page: http://www.ncbi.nlm.nih.gov/sites/gquery/ Query all NCBI databases via any search term: http://www.ncbi.nlm.nih.gov/sites/gquery?term=TP53	(NCBI, 2014)
EBI Search (EB-eye)	Home page: http://www.ebi.ac.uk/ebisearch/ Query via any search term, e.g. (a) gene symbol (b) Ensembl gene (c) RefSeq nucleotide (d) RefSeq protein (e) UniProt ID: (a) http://www.ebi.ac.uk/ebisearch/search.ebi?db=allebi&query=TP53 (p53) (b) http://www.ebi.ac.uk/ebisearch/search.ebi?db=allebi&query=ENSG00000156970 (c) http://www.ebi.ac.uk/ebisearch/search.ebi?db=allebi&query=NM_004153 (d) http://www.ebi.ac.uk/ebisearch/search.ebi?db=allebi&query=NP_009125 (e) http://www.ebi.ac.uk/ebisearch/search.ebi?db=allebi&query=Q9H410	(Valentin *et al.*, 2010)
Bioinformatic Harvester ***5 species:*** human, mouse, rat, zebrafish, *Arabidopsis*	Home page: http://harvester.kit.edu/ Query across five species via any text term: http://harvester.kit.edu:8080/yacysearch.html?query=TP53 (p53) Query via any gene, transcript or protein ID from the five species, e.g. (a) *Arabidopsis* gene (b) Ensembl gene (c) RefSeq nucleotide (d) UniProt ID: (a) http://harvester.kit.edu:8080/yacysearch.html?query=AT5G24280 (b) http://harvester.kit.edu:8080/yacysearch.html?query=ENSMUSG00000019942 (mouse Cdk1) (c) http://harvester.kit.edu:8080/yacysearch.html?query=NM_004153 (d) http://harvester.kit.edu:8080/yacysearch.html?query=Q9H410 Query via species plus gene term: http://harvester.kit.edu:8080/yacysearch.html?query=mus+musculus+Mcm8 Direct link via IPI IDs for (e) human (f) mouse (g) rat (h) zebrafish (i) Arabidopsis: (e) http://harvester.kit.edu/harvester/human/IPI00968201 (p53) (f) http://harvester.kit.edu/harvester/mouse/IPI00114491 (Cdk1) (g) http://harvester.kit.edu/harvester/rat/IPI00950966 (DNA pol ε) (h) http://harvester.kit.edu/harvester/zebrafish/IPI00890468 (i) http://harvester.kit.edu/harvester/arabidopsis/IPI00516272	(Liebel *et al.*, 2005)

Supplemental Table S18

Resources providing overviews of gene and protein function

Resource	Example web-links	Reference
EBI Gene & Protein Summary ***5 species:*** human, mouse, *S. cerevisiae*, *Drosophila*, *C. elegans*	Direct link via (a) gene symbol (b) *S. cerevisiae* gene (c) *Drosophila* gene (d) *C. elegans* gene (e) Ensembl gene ID (f) RefSeq protein (g) UniProt ID: (a) http://www.ebi.ac.uk/s4/summary/molecular?term=TP53 (p53) (b) http://www.ebi.ac.uk/s4/summary/molecular?term=YBR156C (c) http://www.ebi.ac.uk/s4/summary/molecular?term=FBgn0004106 (d) http://www.ebi.ac.uk/s4/summary/molecular?term=K08E3.6 (e) http://www.ebi.ac.uk/s4/summary/molecular?term=ENSG00000156970 (f) http://www.ebi.ac.uk/s4/summary/molecular?term=NP_035145 (g) http://www.ebi.ac.uk/s4/summary/molecular?term=P04637	(McWilliam *et al.*, 2013)
GeneCards (human genes)	Home page: http://www.genecards.org/ Direct link via human gene symbol: http://www.genecards.org/cgi-bin/carddisp.pl?gene=TP53 (p53) Query via human IDs: (a) Ensembl gene (b) RefSeq nucleotide (c) RefSeq protein (d) UniProt: (a) http://www.genecards.org/index.php?path=/Search/keyword/ENSG00000156970 (b) http://www.genecards.org/index.php?path=/Search/keyword/NM_004153 (c) http://www.genecards.org/index.php?path=/Search/keyword/NP_004144 (d) http://www.genecards.org/index.php?path=/Search/SymbolAliasDescIdentifier/Q9H410	(Stelzer *et al.*, 2011)
neXtProt (human proteins)	Home page: http://www.nextprot.org/ Query via (a) any text term (b) Ensembl gene ID: (a) http://www.nextprot.org/db/search#importin (b) http://www.nextprot.org/db/search#ENSG00000141510 (p53) Direct link via human UniProt ID: http://www.nextprot.org/db/entry/NX_P06493 (Cdk1)	Gaudet *et al.*, 2013)
InterMine family	**FlyMine** home page: http://www.flymine.org/ Direct link via FlyBase ID: http://www.flymine.org/query/portal.do?externalids=FBgn0003525 Query via any term (e.g. gene name, UniProt ID): http://www.flymine.org/release-38.0/keywordSearchResults.do?searchTerm=grapes **YeastMine** (*S. cerevisiae*) home page: http://yeastmine.yeastgenome.org/ Direct link via gene name, SGD ID, systematic name: http://yeastmine.yeastgenome.org/yeastmine/portal.do?externalids=CDC28 Query via any term (e.g. UniProt ID): http://yeastmine.yeastgenome.org/yeastmine/keywordSearchResults.do?searchTerm=P23291 **metabolicMine** (includes *H. sapiens*) home page: http://www.metabolicmine.org/ Direct link via gene symbol, Ensembl ID: http://www.metabolicmine.org/beta/portal.do?externalids=TP53 Query via any term (e.g. UniProt ID): http://www.metabolicmine.org/beta/keywordSearchResults.do?searchTerm=P04637 **MouseMine** home page: http://www.mousemine.org/mousemine/ Direct link via gene symbol, MGI, Ensembl, RefSeq, UniProt ID: http://www.mousemine.org/mousemine/portal.do?externalids=NM_011640 Query via any term: http://www.mousemine.org/mousemine/keywordSearchResults.do?searchTerm=Cyclin **RatMine** home page: http://ratmine.mcw.edu/ Direct link via gene symbol, Ensembl, RefSeq ID: http://ratmine.mcw.edu/ratmine/portal.do?externalids=NM_053677 Query via any term (e.g. UniProt ID): http://ratmine.mcw.edu/ratmine/keywordSearchResults.do?searchTerm=Q63699 **ZebrafishMine** home page: http://zebrafishmine.org/ Direct link via gene name, ZFIN, UniProt ID: http://zebrafishmine.org/portal.do?externalids=ZDB-GENE-000330-5 Query via any term: http://zebrafishmine.org/keywordSearchResults.do?searchTerm=Cyclin **modMine** (fly, worm) home page: http://intermine.modencode.org/ Direct link via WormBase or FlyBase ID: http://intermine.modencode.org/query/portal.do?externalids=WBGene00009368 Query via any term : http://intermine.modencode.org/release-32/keywordSearchResults.do?searchTerm=cdk	(Smith *et al.*, 2012)
canSAR	Home page: https://cansar.icr.ac.uk/ Direct link to "Molecular Target Synopsis Overview" page, via UniProt ID: https://cansar.icr.ac.uk/cansar/molecular-targets/P04637/	(Bulusu *et al.*, 2014)
BioGPS	Home page: http://www.biogps.org/ Query via any term: http://www.biogps.org/#goto=search&query=ABL1 Direct link to gene report via (a) NCBI Gene (b) RefSeq or UniProt ID: (a) http://www.biogps.org/#goto=genereport&id=7157 (b) http://www.biogps.org/#goto=search&query=P04637	(Wu *et al.*, 2013)

Supplemental Table S19

Tools for retrieving information about multiple gene or protein hits

Resource	Web address for input or upload of IDs	References
Annotation of results tables		
PANTHER *See* Supplemental Method S4.	http://www.pantherdb.org/genes/batchIdSearch.jsp - "Batch ID Search"	(Mi *et al.*, 2013)
DAVID *See* Supplemental Method S5.	http://david.abcc.ncifcrf.gov/tools.jsp - "Analysis Wizard"	(Huang da *et al.*, 2009b)
UniProt *See* Supplemental Method S6.	*Via "Retrieve" tab (for UniProt IDs):* http://www.uniprot.org/?tab=batch *Via "ID Mapping" tab (for non-UniProt IDs):* http://www.uniprot.org/?tab=mapping	(UniProt, 2014)
bioCompendium	http://biocompendium.embl.de/	
Graphical display of protein domains		
CDD	http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi - "Batch Web CD-Search Tool". Click "Submit", then "Browse results in graphical format"	(Marchler-Bauer *et al.*, 2013)
SMART	http://smart.embl.de/smart/batch.pl - "precomputed results batch retrieval"	(Letunic *et al.*, 2012)
Interaction network diagram of multiple gene products		
STRING *See* Supplemental Method S7.	http://string-db.org/newstring_cgi/show_input_page.pl?input_page_type=multiple_identifiers - "multiple names" tab	(Franceschini *et al.*, 2013)
Interactome3D	http://interactome3d.irbbarcelona.org/ - "Query interactions with proteins" section	(Mosca *et al.*, 2013)
Cytoscape	*Link for software download, documentation etc:* http://www.cytoscape.org/ *App store (plugins):* http://apps.cytoscape.org/	(Smoot *et al.*, 2011; Saito *et al.*, 2012; Lotia *et al.*, 2013)
Pathway analysis of multiple gene products		
GenMAPP	*Link for software download, documentation etc:* http://www.genmapp.org/	(Salomonis *et al.*, 2007)
Multiple gene reports		
BioGPS	http://www.biogps.org/#goto=welcome - "Search genes here" section	(Wu *et al.*, 2013)
Bioinformatic enrichment tools		
DAVID	http://david.abcc.ncifcrf.gov/tools.jsp - "Analysis Wizard"	(Huang da *et al.*, 2009b)
GeneCodis	http://genecodis.cnb.csic.es/	(Tabas-Madrid *et al.*, 2012)
GenePattern	*Link for software download, documentation etc:* http://www.broadinstitute.org/cancer/software/genepattern/	(Reich *et al.*, 2006)
GO-Elite	http://www.genmapp.org/go_elite/	(Zambon *et al.*, 2012)
GOToolBox	http://genome.crg.es/GOToolBox/	(Martin *et al.*, 2004)
STRING	http://string-db.org/newstring_cgi/show_input_page.pl?input_page_type=multiple_identifiers - "multiple names" tab	(Franceschini *et al.*, 2013)
The above is a very limited selection of the bioinformatic enrichment tools available. For a more comprehensive listing, description and comparison of such software, readers are invited to consult the following reviews, and links & references therein: Hedegaard *et al.*, 2009; Huang da *et al.*, 2009a; Kouskoumvekaki *et al.*, 2013		

Supplemental Table S20

Hyperlinked table: a selection of resources for the example cell proliferation factors

			Gene ontology, gene & protein overview, search engine						Interactions and pathways			
Gene	**UniProt**	**Protein name**	**QuickGO**	**OMIM**	**GeneCards**	**EBI Summary**	**neXtProt**	**Bioinfo Harvester**	**IntAct**	**BioGRID**	**STRING**	**Reactome**
TP53	**P04637**	**Tumor Suppressor p53**	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
CDK1	**P06493**	**Cyclin-dependent kinase 1**	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
POLE	**Q07864**	**DNA Polymerase Epsilon**	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
KPNB1	Q14974	Importin subunit beta-1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
CHEK1	O14757	Ser/Thr protein kinase Chk1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
AURKB	Q96GD4	Aurora kinase B	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
RPA2	P15927	Replication protein A, 32 kDa subunit	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
CDT1	Q9H211	DNA replication factor Cdt1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
MCMBP	Q9BTE3	MCM complex-binding protein	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
TUBG1	P23258	Tubulin gamma-1 chain	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
RAN	P62826	GTP-binding nuclear protein Ran	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
RANGRF	Q9HD47	Ran guanine nucleotide factor Mog1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
BLM	P54132	Bloom syndrome protein	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
PCNA	P12004	Poliferating Cell Nuclear Antigen	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
SETD8	Q9NQR1	Pr-Set7	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
RCC1	P18754	Regulator of chromosome condensation	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
MCM5	P33992	DNA replication factor MCM5	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
CDC25C	P30307	M-phase inducer phosphatase	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
PLK1	P53350	Ser/Thr protein kinase PLK1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome
MZT1	Q08AG7	Mitotic-spindle organizing protein 1	GO	OMIM	GeneCards	EBI Sum	neXtProt	Harvester	IntAct	BioGRID	STRING	Reactome

			Protein domains and features							Protein structure, phosphorylation		Drugs
Gene	**UniProt**	**Protein name**	**Pfam**	**SMART**	**InterPro**	**CDD**	**ELM**	**Dasty**	**Annie**	**PDBsum**	**Phospho Site Plus**	**canSAR**
TP53	**P04637**	**Tumor Suppressor p53**	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
CDK1	**P06493**	**Cyclin-dependent kinase 1**	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
POLE	**Q07864**	**DNA Polymerase Epsilon**	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
KPNB1	Q14974	Importin subunit beta-1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
CHEK1	O14757	Ser/Thr protein kinase Chk1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
AURKB	Q96GD4	Aurora kinase B	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
RPA2	P15927	Replication protein A, 32 kDa subunit	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
CDT1	Q9H211	DNA replication factor Cdt1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
MCMBP	Q9BTE3	MCM complex-binding protein	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
TUBG1	P23258	Tubulin gamma-1 chain	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
RAN	P62826	GTP-binding nuclear protein Ran	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
RANGRF	Q9HD47	Ran guanine nucleotide factor Mog1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
BLM	P54132	Bloom syndrome protein	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
PCNA	P12004	Poliferating Cell Nuclear Antigen	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
SETD8	Q9NQR1	Pr-Set7	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
RCC1	P18754	Regulator of chromosome condensation	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
MCM5	P33992	DNA replication factor MCM5	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
CDC25C	P30307	M-phase inducer phosphatase	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
PLK1	P53350	Ser/Thr protein kinase PLK1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR
MZT1	Q08AG7	Mitotic-spindle organizing protein 1	Pfam	SMART	InterPro	CDD	ELM	Dasty	Annie	PDBsum	PSP	canSAR

Supplemental Method S1

Procedure for creating a hyperlinked results table.

Hyperinks are created in a spreadsheet - based on gene, transcript or protein IDs - linking to specific entries in bioinformatic database resources.

The most commonly-used spreadsheet software (including *Microsoft Excel*, *Google docs*, *Apple Numbers*, *OpenOffice.org Calc*) allows web-links to be added to spreadsheet cells by means of a function called "HYPERLINK".

For example, the following UniProt link: http://www.uniprot.org/uniprot/P04637 can be expressed using a spreadsheet HYPERLINK function, thus:

```
=HYPERLINK("http://www.uniprot.org/uniprot/P04637","UniProt")
```

Using this function, a hyperlink appears within the spreadsheet cell (visible with the text "UniProt"), which when clicked opens that web-page in the user's default web browser.

Expressing the web-link as a HYPERLINK function enables a reference to another spreadsheet cell (e.g. cell "A2") - and thus the contents of that cell - to be incorporated into the hyperlink.

So if cell A2 contains a gene or protein ID (for example the UniProt ID "P04637"), then a HYPERLINK function can be entered in a cell (typically nearby in the same row, e.g. cell B2) to refer to the contents of cell A2, thus:

```
=HYPERLINK("http://www.uniprot.org/uniprot/"&A2,"UniProt")
```

or

```
=HYPERLINK("http://www.uniprot.org/uniprot/"&A2&"#section_features","UniProt-Feat")
```

This HYPERLINK function can then be copied to all cells in that column (with the relevant cell references being updated) by "dragging down" the cell to the bottom of the table. The addition of multiple hyperlinks to results tables can be automated using a macro, where available.

Supplemental Method S2

Procedure for extracting gene symbols from UniProt protein entries.

This allows direct access to resources based on gene symbols, where the primary hits are UniProt proteins.

In UniProt, the standard format of a header row contains the corresponding official gene symbol, proceeded by the text "GN=", such as in the example below:

>sp|Q9H1A4|APC1_HUMAN Anaphase-promoting complex subunit 1 OS=Homo sapiens GN=ANAPC1 PE=1 SV=1

Extracting the gene symbol "ANAPC1" from the header row is a text-trimming exercise, easily performed in standard spreadsheet software using two search-and-replace operations.

To trim the surrounding text to just the gene symbol (e.g. "ANAPC1"):

1. Create a new column next to that containing the protein headings.
2. Select the column containing the protein headings, copy the contents to the clipboard, then paste the contents into the newly-created column.
3. In the new column, search for "*GN=", and replace with "" (nothing).
4. In the same column, search for " *", and replace with "" (nothing).

The new column will now contain only the gene symbols, ready for linking to gene-oriented databases.

Supplemental Method S3

Procedure for obtaining reviewed UniProt (i.e. Swiss-Prot) IDs from a list of gene symbols.

This allows resources based on UniProt protein IDs to be accessed in cases where only the gene symbols are known.

1. Obtaining a list of gene IDs from gene names/symbols

Go to the relevant organism-specific gene database website (e.g. HGNC, MGI).

Use the database's facility to batch identify gene IDs from gene names / symbols.

Select this list of gene IDs (command + drag for an HTML table), and copy to the clipboard.

2. Converting Gene IDs to UniProt IDs

Go to the UniProt's "ID Mapping" site: http://www.uniprot.org/?tab=mapping.

In the "Database identifiers" window, paste in the gene IDs. In the "From" pull-down menu, choose the appropriate organism-specific gene database. The "To" option should then read "UniProtKB AC".

Click "Map": this generates a list of input IDs and their mapped UniProt counterparts.

Click "UniProtKB": this produces a table of the mapped UniProt entries, with some attributes.

Click "Show only reviewed" (where appropriate): this shows only the Swiss-Prot entries.

Click "Customize", choose "Gene names" then "Show »". Check the "Name" box, and click "Show »". Click "Save": the UniProt table should then be updated, with an extra column headed "Gene names (PREFERRED)".

The contents of the columns "Entry" (UniProt ID) and "Gene Names (PREFERRED)" can then be copied from the UniProt table, and pasted back into the original results spreadsheet. Note that the contents of these two new columns may need to be sorted so that they match the order in the original hits table.

Supplemental Method S4

Procedure for generating an annotated results table (using PANTHER).

Annotations include: GO-slim terms, and PANTHER "protein class" and "Pathway" terms.

1. Importing hits from the results table into PANTHER

In results spreadsheet: copy the list of IDs from the results table.

In a web browser: go to the PANTHER website: http://www.pantherdb.org/

Under the "Gene List Analysis" tab, paste the copied IDs into the box headed "Enter ids and or select file for batch upload". Select the relevant organism from the list.

Select "Functional classification viewed in gene list". Click "submit".

A web-based annotation table will be presented.

To generate pie-charts of GO- and PANTHER-term usage, click the pie-chart icon.

2. Transferring the PANTHER functional annotation table into a spreadsheet

Above the table, after "Send list to:", select "File". Save the file (default filename "pantherGeneList.txt") to a known location.

In spreadsheet program: import the PANTHER text file. (For *Microsoft Excel*: open a new spreadsheet. Choose >File >Import…. Choose "Text file", click "Import". Select the file saved by PANTHER, and click "Get Data". In the Text Import Wizard, choose "Delimited", then "Next >". Under "Delimiters", choose "Tab", then "Next >". Click "Finish". In "Import Data", click "OK".)

Output: columns appear (without header names) in the following order:

A : Gene ID (PANTHER format)
B : Mapped IDs
C : Name; Gene Symbol; Orthologue
D : PANTHER Family/Subfamily
E : GO Molecular Function ("PANTHER slim")
F : GO Biological Process ("PANTHER slim")
G : GO Cellular Component ("PANTHER slim")
H : PANTHER Protein Class
I : PANTHER Pathway
J : Species

Note: the rows in the PANTHER-annotated table may need to be sorted so that they match those in the original hits table, before being integrated into the results spreadsheet.

Supplemental Method S5

Procedure for generating an annotated results table (using DAVID).

Annotations include: full GO and OMIM terms, BioCarta and KEGG pathways, and InterPro and SMART domains.

1. Importing hits from the results table into *DAVID*

In results spreadsheet: copy the list of IDs from the results table.

In a web browser: go to the DAVID Analysis Wizard: http://david.abcc.ncifcrf.gov/tools.jsp

Click the "Upload" tab (if not already selected).

Under "Step 1: Enter Gene List" and "A: Paste a list", paste in the list of IDs.

Under "Step 2: Select Identifier", choose the type of ID from the pull-down menu.

Under "Step 3: List Type", select "Gene List".

Under "Step 4: Submit List", click "Submit List".

2. Generating a *DAVID* "Functional Annotation Table"

Under "Step 2. Analyze above gene list with one of DAVID tools", click "Functional Annotation Table".

In the "Annotation Summary Results" page, under "Combined View for Selected Annotation", click "Functional Annotation Table". This opens a web-based table with multiple functional annotations.

3. Transferring the *DAVID* Functional Annotation Table into a spreadsheet

In DAVID "Functional Annotation Table" page: click "Download File" (upper-right corner of the table): this opens the annotations in a plain-text format in a browser window.

Select all of the text, copy to the clipboard.

In spreadsheet program: open a new blank spreadsheet document. Paste the clipboard contents into the blank spreadsheet (note- for *Microsoft Excel*, use "Paste Special", choosing "Text"). The list of proteins and annotations will appear in separate columns.

Note: the rows in the DAVID-annotated table may need to be sorted so that they match those in the original hits table, before being integrated into the results spreadsheet.

Supplemental Method S6

Procedure for generating an annotated results table (using UniProt).

Annotations include: amino acid sequence, disease association, domain descriptions, features, gene symbol & synonyms, general annotation, GO terms (combined), InterPro classification, keywords, protein names, subcellular location.

1. Importing hits from the results table into UniProt

In the results spreadsheet: select the IDs of the hits in the table, copy to the clipboard.

For hits in UniProt format: use the "Retrieve" function: http://www.uniprot.org/?tab=batch. Paste the IDs into the window headed "UniProt identifiers". Click "Retrieve".

For hits in other formats: use the "ID Mapping" function: http://www.uniprot.org/?tab=mapping. Paste the IDs into the window headed "Database identifiers". In the "From" drop-down menu choose the database corresponding to the ID type. Click "Map".

The hits then appear (as proteins) in a web-based table in UniProt.

2. Choosing annotation attributes to be added

Above the UniProt table, click "Customize".

Under the "Columns" list appears a list of features that can be added to annotate the hits table. Select the desired features and click "Show »".

Click "Save" to update the UniProt table with annotations of the chosen type.

3. Transferring the UniProt-annotated table into a spreadsheet

Click the "Download" button.

To open in the table in *Microsoft Excel*, under the "Excel" heading, click "Open".

The table can be saved in the tab-delimited format for import into other spreadsheet programs: under the "Tab-Delimited" heading, click "Open". Select the text, copy to clipboard. Paste into the program of your choice.

Note: the rows in the UniProt-annotated table may need to be sorted so that they match those in the original hits table, before being integrated into the results spreadsheet.

Supplemental Method S7

Procedure for generating a network diagram (using STRING).

1. ***In the results spreadsheet:*** select the list of gene symbols or UniProt IDs, and copy these to the clipboard.

2. ***In a web browser:*** go to the STRING "multiple names" tab. Web-link:
   http://string-db.org/newstring_cgi/show_input_page.pl?&input_page_type=multiple_identifiers
   In the box headed "list of names", paste in the gene symbols or UniProt IDs. Click "GO!".
   A list of protein names matched to the IDs will be presented. Click "Continue ->".

A STRING network diagram should then be displayed. The default view is the "evidence view", i.e. with interactions (edges) colored by interaction type. From here, various alternative view options can be explored, and the positions of the proteins can be adjusted. The "save" button allows the diagram to be saved as a high-resolution graphic.

Supplemental References

Alcántara, R., *et al.* (2013). The EBI enzyme portal. Nucleic Acids Res *41*, D773-780.

Amberger, J., Bocchini, C., and Hamosh, A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM). Human mutation *32*, 564-567.

Arnaud, M.B., *et al.* (2012). The Aspergillus Genome Database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. Nucleic Acids Res *40*, D653-659.

Asplund, A., Edqvist, P.H., Schwenk, J.M., and Ponten, F. (2012). Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. Proteomics *12*, 2067-2077.

Bao, Y., Federhen, S., Leipe, D., Pham, V., Resenchuk, S., Rozanov, M., Tatusov, R., and Tatusova, T. (2004). National center for biotechnology information viral genomes project. J Virol *78*, 7291-7298.

Barrett, T., *et al.* (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res *41*, D991-995.

Basu, S., Fey, P., Pandit, Y., Dodson, R., Kibbe, W.A., and Chisholm, R.L. (2013). dictyBase 2013: integrating multiple Dictyostelid species. Nucleic Acids Res *41*, D676-683.

Becker, K.G., Barnes, K.C., Bright, T.J., and Wang, S.A. (2004). The genetic association database. Nat Genet *36*, 431-432.

Bento, A.P., *et al.* (2014). The ChEMBL bioactivity database: an update. Nucleic Acids Res *42*, D1083-1090.

Blake, J.A. (2013). Gene Ontology Annotations and Resources. Nucleic Acids Res *41*, D530-D535.

Boeker, M., Vach, W., and Motschall, E. (2013). Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Med Res Methodol *13*, 131.

Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P., and Swaminathan, G.J. (2014). DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res *42*, D993-D1000.

Bulusu, K.C., Tym, J.E., Coker, E.A., Schierz, A.C., and Al-Lazikani, B. (2014). canSAR: updated cancer research and drug discovery knowledgebase. Nucleic Acids Res *42*, D1040-1047.

Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O.n., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011). Pathway Commons, a web resource for biological pathway data. Nucleic acids research *39*, D685-690.

Chatr-Aryamontri, A., *et al.* (2013). The BioGRID interaction database: 2013 update. Nucleic Acids Res *41*, D816-823.

Chernorudskiy, A.L., Garcia, A., Eremin, E.V., Shorina, A.S., Kondratieva, E.V., and Gainullin, M.R. (2007). UbiProt: a database of ubiquitylated proteins. BMC Bioinformatics *8*, 126.

Costanzo, M.C., *et al.* (2014). Saccharomyces genome database provides new regulation data. Nucleic Acids Res *42*, D717-725.

Croft, D.*, et al.* (2014). The Reactome pathway knowledgebase. Nucleic Acids Res *42*, D472-477.

Davis, A.P.*, et al.* (2013). The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res *41*, D1104-1114.

de Beer, T.A., Berka, K., Thornton, J.M., and Laskowski, R.A. (2014). PDBsum additions. Nucleic Acids Res *42*, D292-296.

Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. (2011). Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Res *39*, D261-267.

Dinkel, H.*, et al.* (2014). The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res *42*, D259-266.

Drabkin, H.J., and Blake, J.A. (2012). Manual Gene Ontology annotation workflow at the Mouse Genome Informatics Database. Database (Oxford) *2012*, bas045.

Fernandez, J.M., Hoffmann, R., and Valencia, A. (2007). iHOP web services. Nucleic Acids Res *35*, W21-26.

Finn, R.D.*, et al.* (2014). Pfam: the protein families database. Nucleic Acids Res *42*, D222-230.

Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F., and Apweiler, R. (2004). IntEnz, the integrated relational enzyme database. Nucleic Acids Res *32*, D434-437.

Flicek, P.*, et al.* (2014). Ensembl 2014. Nucleic Acids Res *42*, D749-755.

Forbes, S.A.*, et al.* (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res *39*, D945-950.

Franceschini, A.*, et al.* (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res *41*, D808-815.

Gaudet, P.*, et al.* (2013). neXtProt: Organizing Protein Knowledge in the Context of Human Proteome Projects. J Proteome Res *12*, 293-298.

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S.H. (2010). The NCBI BioSystems database. Nucleic Acids Res *38*, D492-496.

Giustini, D., and Barsky, E. (2005). A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations. J. Canad. Health Libr. Assoc. *26*, 85-89.

Gnad, F., Gunawardena, J., and Mann, M. (2011). PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res *39*, D253-260.

Good, B.M., Clarke, E.L., de Alfaro, L., and Su, A.I. (2012). The Gene Wiki in 2011: community intelligence applied to human gene annotation. Nucleic Acids Res *40*, D1255-1261.

Gray, K.A., Daugherty, L.C., Gordon, S.M., Seal, R.L., Wright, M.W., and Bruford, E.A. (2013). Genenames.org: the HGNC resources in 2013. Nucleic Acids Res *41*, D545-552.

Gutmanas, A.*, et al.* (2014). PDBe: Protein Data Bank in Europe. Nucleic Acids Res *42*, D285-291.

Harris, T.W.*, et al.* (2014). WormBase 2014: new views of curated biology. Nucleic Acids Res *42*, D789-793.

Hartshorn, M.J. (2002). AstexViewer: a visualisation aid for structure-based drug design. J Comput Aided Mol Des *16*, 871-881.

Hayles, J., Wood, V., Jeffery, L., Hoe, K.L., Kim, D.U., Park, H.O., Salas-Pino, S., Heichinger, C., and Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. Open Biol *3*, 130053.

He, Y.*, et al.* (2012). dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. Nucleic Acids Res *40*, D964-971.

Hedegaard, J.*, et al.* (2009). Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. BMC Proc *3 Suppl 4*, S5.

Herraez, A. (2006). Biomolecules in the computer: Jmol to the rescue. Biochem Mol Biol Educ *34*, 255-261.

Hibbs, M.A., Hess, D.C., Myers, C.L., Huttenhower, C., Li, K., and Troyanskaya, O.G. (2007). Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics *23*, 2692-2699.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res *40*, D261-270.

Howe, D.G.*, et al.* (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res *41*, D854-860.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Hunter, S.*, et al.* (2012). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res *40*, D306-D312.

Huntley, R.P., Binns, D., Dimmer, E., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGO: a user tutorial for the web-based Gene Ontology browser. Database (Oxford) *2009*, bap010.

Hutchins, J.R.A.*, et al.* (2010). Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science *328*, 593-599.

James-Zorn, C.*, et al.* (2013). Xenbase: expansion and updates of the Xenopus model organism database. Nucleic Acids Res *41*, D865-870.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res *42*, D199-205.

Kapushesky, M.*, et al.* (2012). Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments. Nucleic Acids Res *40*, D1077-1081.

Karolchik, D.*, et al.* (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res *42*, D764-770.

Kelder, T., van Iersel, M.P., Hanspers, K., Kutmon, M., Conklin, B.R., Evelo, C.T., and Pico, A.R. (2012). WikiPathways: building research communities on biological pathways. Nucleic Acids Res *40*, D1301-1307.

Kerrien, S.*, et al.* (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res *40*, D841-846.

Kersey, P.J.*, et al.* (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res *42*, D546-552.

Kersey, P.J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004). The International Protein Index: an integrated database for proteomics experiments. Proteomics *4*, 1985-1988.

Kim, D.U.*, et al.* (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol *28*, 617-623.

Kouskoumvekaki, I., Shublaq, N., and Brunak, S. (2013). Facilitating the use of large-scale biological data and tools in the era of translational bioinformatics. Brief Bioinform.

Lamesch, P.*, et al.* (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res *40*, D1202-1210.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res *42*, D980-985.

Laulederkind, S.J.*, et al.* (2012). Ontology searching and browsing at the Rat Genome Database. Database (Oxford) *2012*, bas016.

Law, V.*, et al.* (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res *42*, D1091-1097.

Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana, R., Antonescu, V., Chan, a., Cheung, F., and Quackenbush, J. (2005). The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic acids research *33*, D71-74.

Lees, J.G., Lee, D., Studer, R.A., Dawson, N.L., Sillitoe, I., Das, S., Yeats, C., Dessailly, B.H., Rentzsch, R., and Orengo, C.A. (2014). Gene3D: Multi-domain annotations for protein sequence and comparative genome analysis. Nucleic Acids Res *42*, D240-245.

Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res *40*, D302-305.

Liebel, U., Kindler, B., and Pepperkok, R. (2005). Bioinformatic "Harvester": a search engine for genome-wide human, mouse, and rat protein resources. Methods Enzymol *404*, 19-26.

Logan-Klumpler, F.J*., et al.* (2012). GeneDB--an annotation database for pathogens. Nucleic Acids Res *40*, D98-108.

Lotia, S., Montojo, J., Dong, Y., Bader, G.D., and Pico, A.R. (2013). Cytoscape app store. Bioinformatics *29*, 1350-1351.

Lu, C.T., Huang, K.Y., Su, M.G., Lee, T.Y., Bretana, N.A., Chang, W.C., Chen, Y.J., and Huang, H.D. (2013). dbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res *41*, D295-305.

Madej, T*., et al.* (2012). MMDB: 3D structures and macromolecular interactions. Nucleic Acids Res *40*, D461-464.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2011). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res *39*, D52-57.

Marchler-Bauer, A*., et al.* (2013). CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res *41*, D348-352.

Martin, D., Brun, C., Remy, E., Mouren, P., Thieffry, D., and Jacq, B. (2004). GOToolBox: functional analysis of gene datasets based on Gene Ontology. Genome Biol *5*, R101.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res *41*, W597-600.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res *41*, D377-386.

Mi, T*., et al.* (2012). Minimotif Miner 3.0: database expansion and significantly improved reduction of false-positive predictions from consensus sequences. Nucleic Acids Res *40*, D252-260.

Mosca, R., Ceol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein networks. Nat Methods *10*, 47-53.

Müller, H.M., Kenny, E.E., and Sternberg, P.W. (2004). Textpresso: an ontology-based information retrieval and extraction system for biological literature. PLoS Biol *2*, e309.

NCBI. (2014). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res *42*, D7-D17.

Neumann, B*., et al.* (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature *464*, 721-727.

Nishimura, D. (2001). A View From the Web: BioCarta. Biotech Software & Internet Report *2*, 117-120.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res *31*, 3635-3641.

Oldfield, T.J. (2004). A Java applet for multiple linked visualization of protein structure and sequence. J Comput Aided Mol Des *18*, 225-234.

Ooi, H.S., Kwo, C.Y., Wildpaner, M., Sirota, F.L., Eisenhaber, B., Maurer-Stroh, S., Wong, W.C., Schleiffer, A., Eisenhaber, F., and Schneider, G. (2009). ANNIE: integrated de novo protein sequence annotation. Nucleic Acids Res *37*, W435-440.

Orchard, S.*, et al.* (2012). Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods *9*, 345-350.

Pickett, B.E.*, et al.* (2012). ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res *40*, D593-598.

Reardon, S. (2013). Project ranks billions of drug interactions. Nature *503*, 449-450.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat Genet *38*, 500-501.

Rose, P.W.*, et al.* (2013). The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res *41*, D475-D482.

Rustici, G.*, et al.* (2013). ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res *41*, D987-990.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. Nat Methods *9*, 1069-1076.

Salomonis, N., Hanspers, K., Zambon, A.C., Vranizan, K., Lawlor, S.C., Dahlquist, K.D., Doniger, S.W., Stuart, J., Conklin, B.R., and Pico, A.R. (2007). GenMAPP 2: new features and resources for pathway analysis. BMC Bioinformatics *8*, 217.

Schomburg, I.*, et al.* (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res *41*, D764-772.

Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M., and Bateman, A. (2014). TreeFam v9: a new website, more species and orthology-on-the-fly. Nucleic Acids Res *42*, D922-925.

Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res *41*, D344-347.

Smith, R.N.*, et al.* (2012). InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. Bioinformatics *28*, 3163-3165.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics (Oxford, England) *27*, 431-432.

Sönnichsen, B.*, et al.* (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.

St Pierre, S.E., Ponting, L., Stefancsik, R., and McQuilton, P. (2014). FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res *42*, D780-788.

Stelzer, G.*, et al.* (2011). In-silico human genomics with GeneCards. Hum Genomics *5*, 709-717.

Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C.H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics *23*, 1282-1288.

Tabas-Madrid, D., Nogales-Cadenas, R., and Pascual-Montano, A. (2012). GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res *40*, W478-483.

Tassy, O.*, et al.* (2010). The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. Genome Res *20*, 1459-1468.

UniProt. (2014). Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res *42*, D191-198.

Valentin, F., Squizzato, S., Goujon, M., McWilliam, H., Paern, J., and Lopez, R. (2010). Fast and efficient searching of biological data resources--using EB-eye. Brief Bioinform *11*, 375-384.

Villaveces, J.M., Jimenez, R.C., Garcia, L.J., Salazar, G.A., Gel, B., Mulder, N., Martin, M., Garcia, A., and Hermjakob, H. (2011). Dasty3, a WEB framework for DAS. Bioinformatics *27*, 2616-2617.

Wang, Y., Geer, L.Y., Chappey, C., Kans, J.A., and Bryant, S.H. (2000). Cn3D: sequence and structure views for Entrez. Trends Biochem Sci *25*, 300-302.

Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., Shoemaker, B.A., Gindulyte, A., and Bryant, S.H. (2014). PubChem BioAssay: 2014 update. Nucleic Acids Res *42*, D1075-1082.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., and Bryant, S.H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res *37*, W623-633.

Wolfsberg, T.G., Wetterstrand, K.A., Guyer, M.S., Collins, F.S., and Baxevanis, A.D. (2002). A user's guide to the human genome. Nat Genet *32 Suppl*, 1-79.

Wood, V.*, et al.* (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res *40*, D695-D699.

Wu, C., Macleod, I., and Su, A.I. (2013). BioGPS and MyGene.info: organizing online, gene-centric information. Nucleic Acids Res *41*, D561-565.

Zambon, A.C., Gaj, S., Ho, I., Hanspers, K., Vranizan, K., Evelo, C.T., Conklin, B.R., Pico, A.R., and Salomonis, N. (2012). GO-Elite: a flexible solution for pathway and ontology over-representation. Bioinformatics *28*, 2209-2210.