

SOM table legends

Supplemental Online Material for Manning et al (Science 298:1912-1934)

Table S1: Sequences and Classification of Human Kinases

This table lists the classification and sequences for all human protein kinase genes and pseudogenes. Internal accession numbers (SK###) are included to track genes through changes in names and for reference to the accompanying database at <http://www.kinase.com>. Second kinase domains of dual-domain kinases are also included. Pseudogenes are indicated by a 'ps' suffix in the name and a Y in the Pseudogene column. Two putative pseudogenes which contain full open reading frames are designated by the '-rs' suffix and R in the Pseudogene column. The 'Novelty' column estimates the novelty of each gene, as to whether it has been described as a kinase in the literature or via a sequence database record (RefSeq, Genbank or SwissProt). Many genes listed as annotated through database records have also been published in the literature. Genes marked as novel either do not have published human sequences, or are not annotated as kinases.

Table S2: Chromosomal mapping and disease linkage of protein kinases and kinase pseudogenes

Protein kinase genes and pseudogenes were chromosomally mapped using a combination of methods: where sequences overlapped with predicted proteins from the Celera or public genome projects, the computed locations of those predicted ORFs were used. This was combined with data from the literature for several experimentally mapped genes. Where these methods failed or gave conflicting results, sequences were aligned to the public genomic assembly using BLAT (<http://genome.ucsc.edu>). A consensus map location gives a range when the fine mapping from different methods disagreed. Chromosomal map locations were linked to disease loci using OMIM (<http://www.ncbi.nlm.nih.gov/omim/>), and to cancer amplicons using data from S. Knuutila *et al.*, *Am J Pathol* **152**, 1107-23 (1998) and other literature references.

Table S3: Closely related co-mapping kinases

The chromosomal distribution of kinase genes overall is similar to that of total gene count, but several small clusters of 2-3 kinases genes are seen, in which all members of the cluster are from the same family or subfamily. The table lists these clusters as pairs of co-mapping genes. Kinase names in bold are present in multiple co-mapping pairs. Most of these genes belong to families that are expanded more than typical in vertebrates, when compared with invertebrates (worm and fly kinomes), and several pairs come from just a few families (listed in bold) - 10 Eph, 7 NEK, 7 STKR, 6 Ste7 and 6 Src family members. In two cases, the genome contains two pairs of closely mapping genes, where one pair may have derived from a duplication of another pair - p38b/p38g may be a duplication of the p38a/p38d locus, and KIT/PDGFRa is probably a duplication of the FMS/PDGFRb locus.

Table S4: Supplementary data on kinase pseudogenes

Name, classification and sequences of pseudogenes are as in table S1, along with the number of introns (if any), the conservation of canonical splice sites (GT/AG) in those introns, the number of overlapping ESTs and cDNAs for expressed pseudogenes, and notes. Genomic DNA sequences are edited by insertion of “NNN” trinucleotides to correct frameshifts that disrupt the ORF. These appear as single “X” residues within the corresponding protein sequences.

Table S5: Presence of kinase catalytic motifs

Three motifs within the catalytic domain are thought to be critical for catalytic function, each of which contains an almost invariant residue believed to participate in catalysis:

Motif	Notes
VAIK	K interacts with the alpha and beta phosphates of ATP, anchoring and orienting the ATP.
HRD	D is likely to be catalytic, acting as a base acceptor
DFG	D chelates Mg ⁺⁺ ions of ATP

(functional notes from Hardie & Hanks (1995) The Protein Kinase Facts Book (Academic Press)). This table shows the residues present in each of these motifs. To generate this table, each kinase was aligned to the ePK HMM profile to locate the relevant motifs. Kinases which failed to show a canonical motif were reviewed manually, and by alignment with homologs to check if an alternative motif was present. PIKKs were aligned to a PI3-kinase HMM. The presence or absence of these residues does not always correlate with catalytic activity; some instances are noted in the 'notes' column. The HMM P scores are listed for inactive kinases and ‘weak’ kinases with poor scores (< 1e-30).

Table S6: Evolutionary Distribution of Kinase Families

All protein kinase domains from the kinomes of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and human were classified into a system of groups, families and sub-families. Families and subfamilies labeled as “Unique” and “Unclassified” contain divergent genes with no strong similarity to each other. Genes with dual kinase domains are represented twice, second domains are classified with the suffix ‘b’ (e.g. RSK, RSKb). Atypical protein kinases are classified by whole-gene sequence similarity.

Table 7: Additional Domains in Protein Kinases

Numbers indicate the number of genes containing at least one copy of a domain (# genes) and the total number of occurrences of the domain within the kinome (# domains). Further information on each Pfam domain is available at <http://pfam.wustl.edu>. Function class is assigned from Pfam descriptions and the literature.