**File S2. The construction procedures of databases in sequence alignment-based GO prediction methods**

**A. The construction procedures of Gene-GOA**

First, we download all genes with GO annotation from National Center for Biotechnology Information [1] (NCBI). Following the CAFA experiments [2,3], we only select the genes annotated by at least one of the eight experimental evidence codes, including EXP, IDA, IPI, IMP, IGI, IEP, TAS, and IC. Moreover, to explicitly consider the hierarchical structure of GO terms, if a child term is annotated to a gene, all its direct and indirect parents, as defined by the “is\_a” relation in gene ontology database [4] (http://geneontology.org/), are also annotated. The numbers of genes annotated with GO terms for MF, BP, and CC are 40160, 63543, and 55448, respectively, in Gene-GOA.

**B. The construction procedures of genetic sequence database with GO annotation**

To construct a genetic sequence database with GO annotation (GSD-GOA), the RNA sequences of all genes in Gene-GOA are extracted from NCBI [1]. If there is no available RNA sequence for a gene, its genomic DNA sequence is selected. In addition, we discard a few genes which have no available RNA or genomic DNA sequences in NCBI. After this, GSD-GOA includes 39179 sequences with MF terms, 61699 sequences with BP terms, and 54117 sequences with CC terms.

**C. The construction procedures of protein sequence database**

The protein sequence database (PSD) is constructed as follows. For each gene in Gene-GOA, we map it as the corresponding coding protein sequences in UniProt database [5] using a gene-protein mapping table. After this, 78170 genes can be mapped as 119876 protein sequences.

[1] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res 2010;39:D38–51.

[2] Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. Nat Methods 2013;10:221-7.

[3] Jiang Y, Oron TR, Clark WT, Bankapur AR, D’Andrea D, Lepore R, et al. An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Genome Biol 2016;17:1-19.

[4] Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258-61.

[5] Uniprot Consortium. UniProt: a hub for protein information. Nucleic Acids Res 2015;43:D204–12.