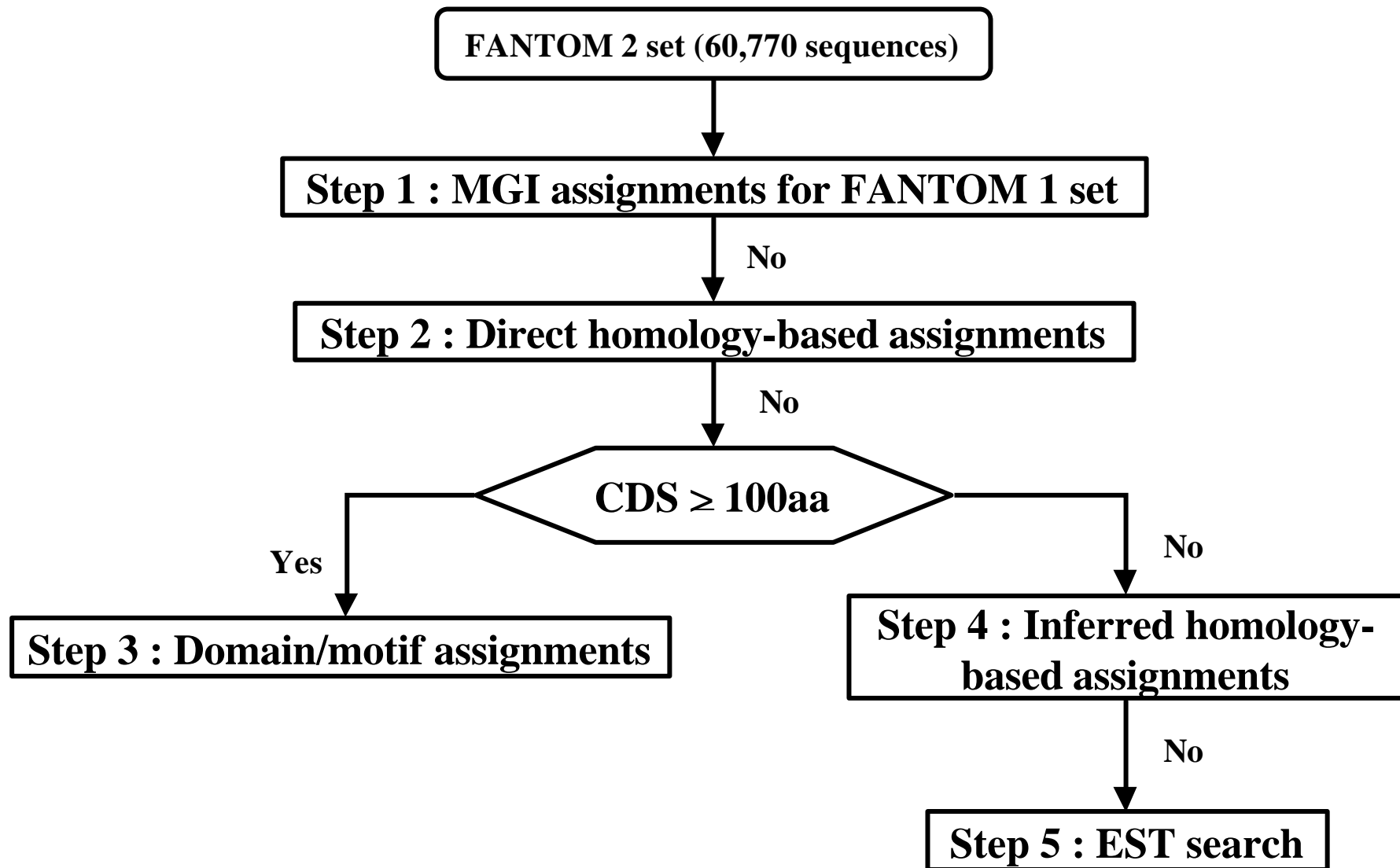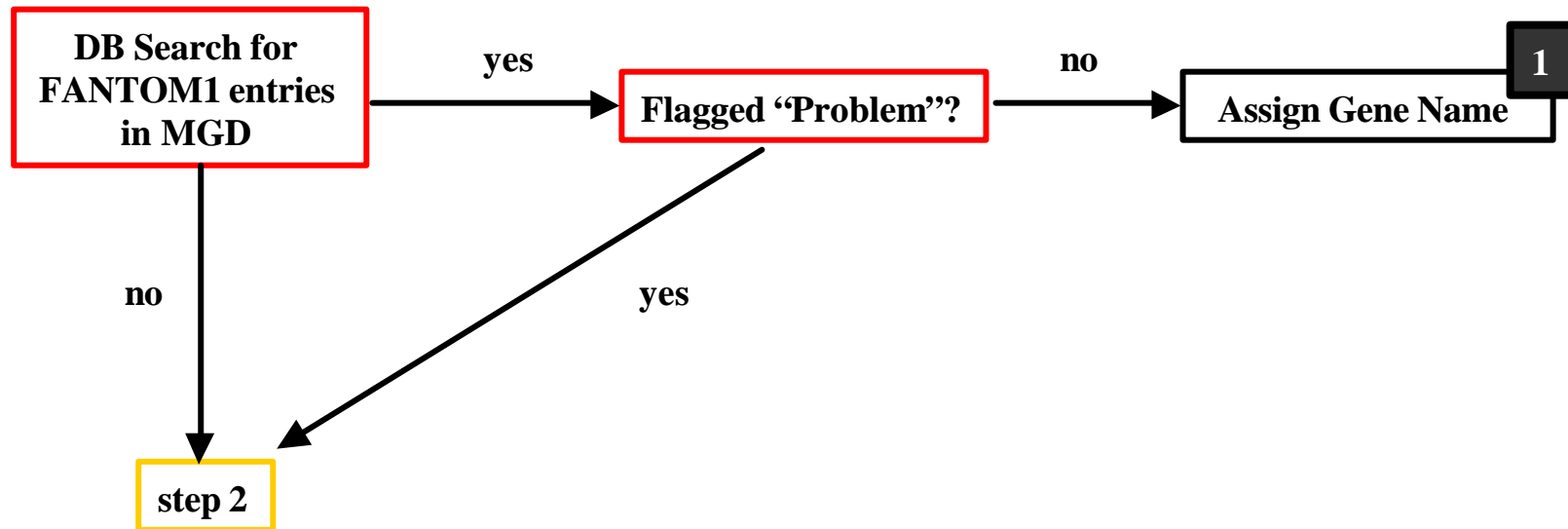Supplementary Information 6A

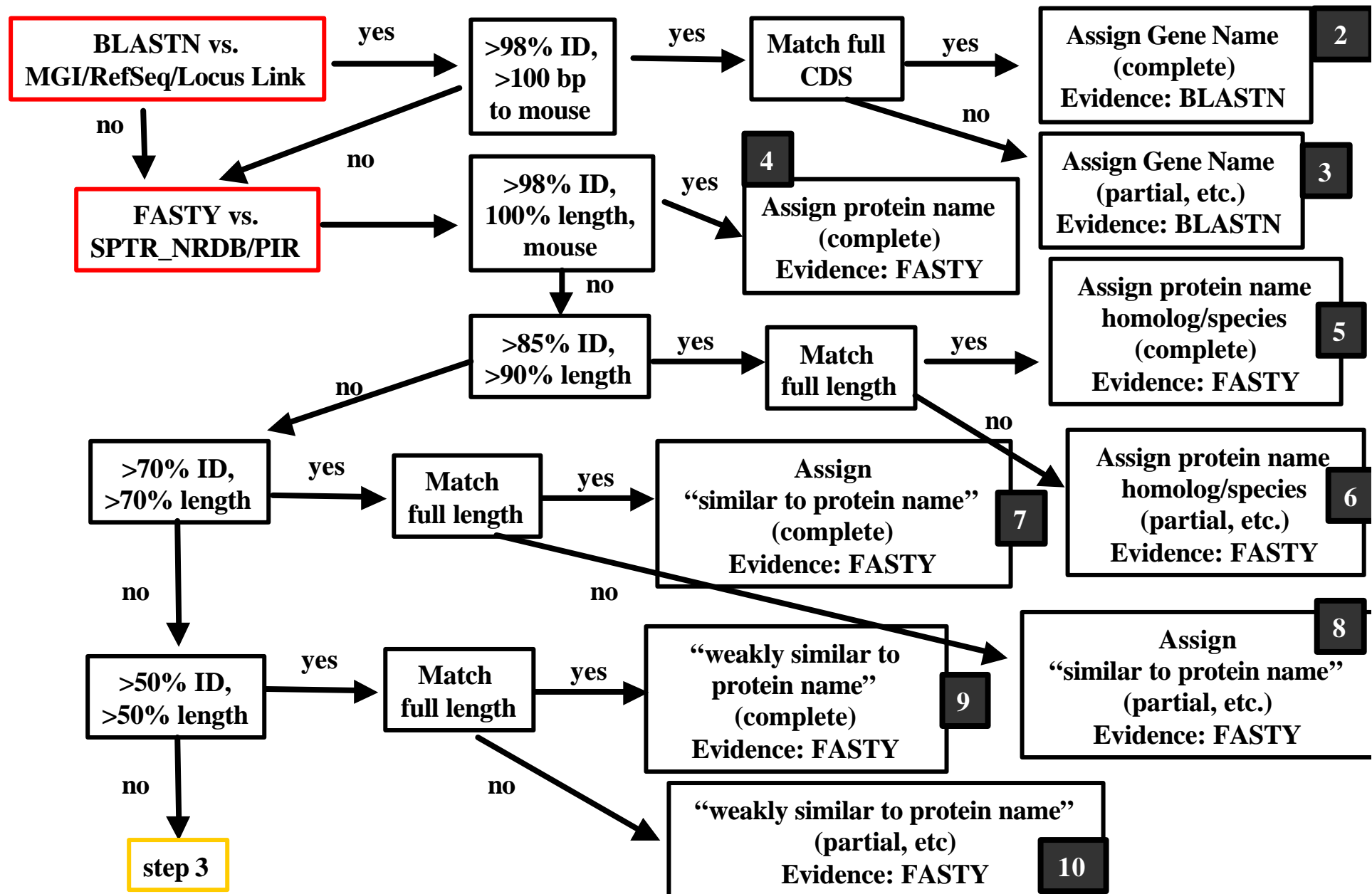## Step 1: MGI Assignments for FANTOM 1 set

Supplementary Information 6C

Step 2: Direct Homology-based Assignments

Supplementary Information 6D
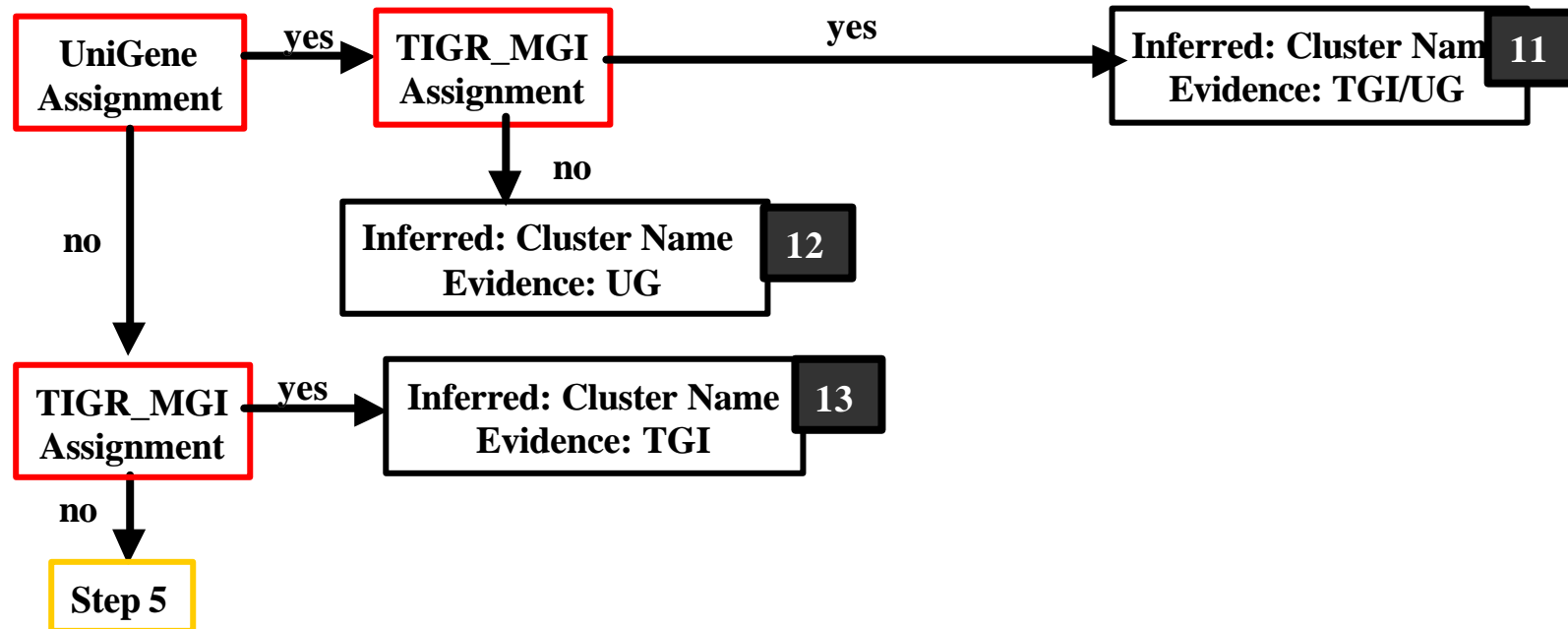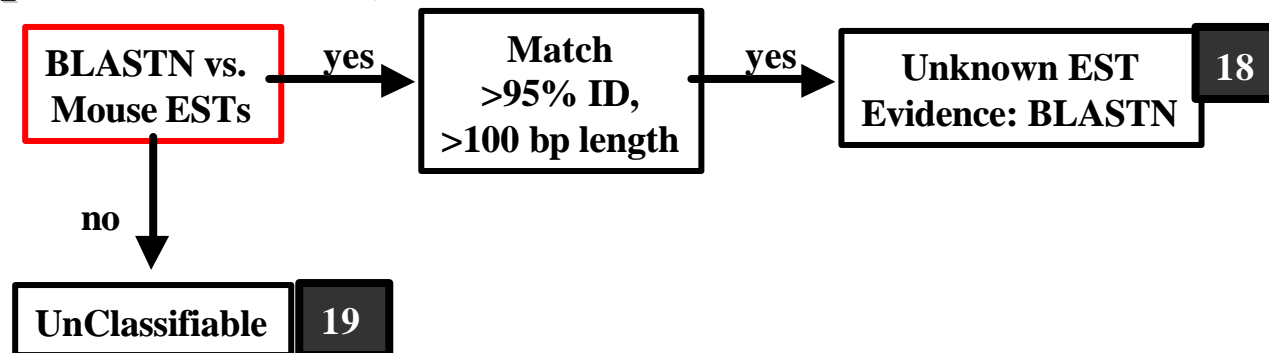
Step 3: Domain/Motif Assignments

```
                    CDS ──no──▶ Step 4
                     │
                    yes
                     │
                     ▼
              InterPro ──yes──▶  ┌──────────────────────────────────┐  14
              Assignment         │ Assign InterPro Annotation:      │
                     │           │ hypothetical XX containing protein│
                    no           │ Evidence: InterPro               │
                     │           └──────────────────────────────────┘
                     ▼
                MDS ──yes──▶  ┌──────────────────────────────────┐  15
              Assignment      │ Assign MDS Annotation:           │
                     │        │ hypothetical XX containing protein│
                    no        │ Evidence: MDS                    │
                     │        └──────────────────────────────────┘
                     ▼
               SCOP ──yes──▶  ┌──────────────────────────────────────┐  16
                     │        │ Assign SCOP Annotation:              │
                    no        │ hypothetical XX structure containing protein│
                     │        │ Evidence: SCOP                       │
                     ▼        └──────────────────────────────────────┘
           ┌──────────────────────┐  17
           │ hypothetical protein │
           │ Evidence: PROCREST   │
           └──────────────────────┘
```
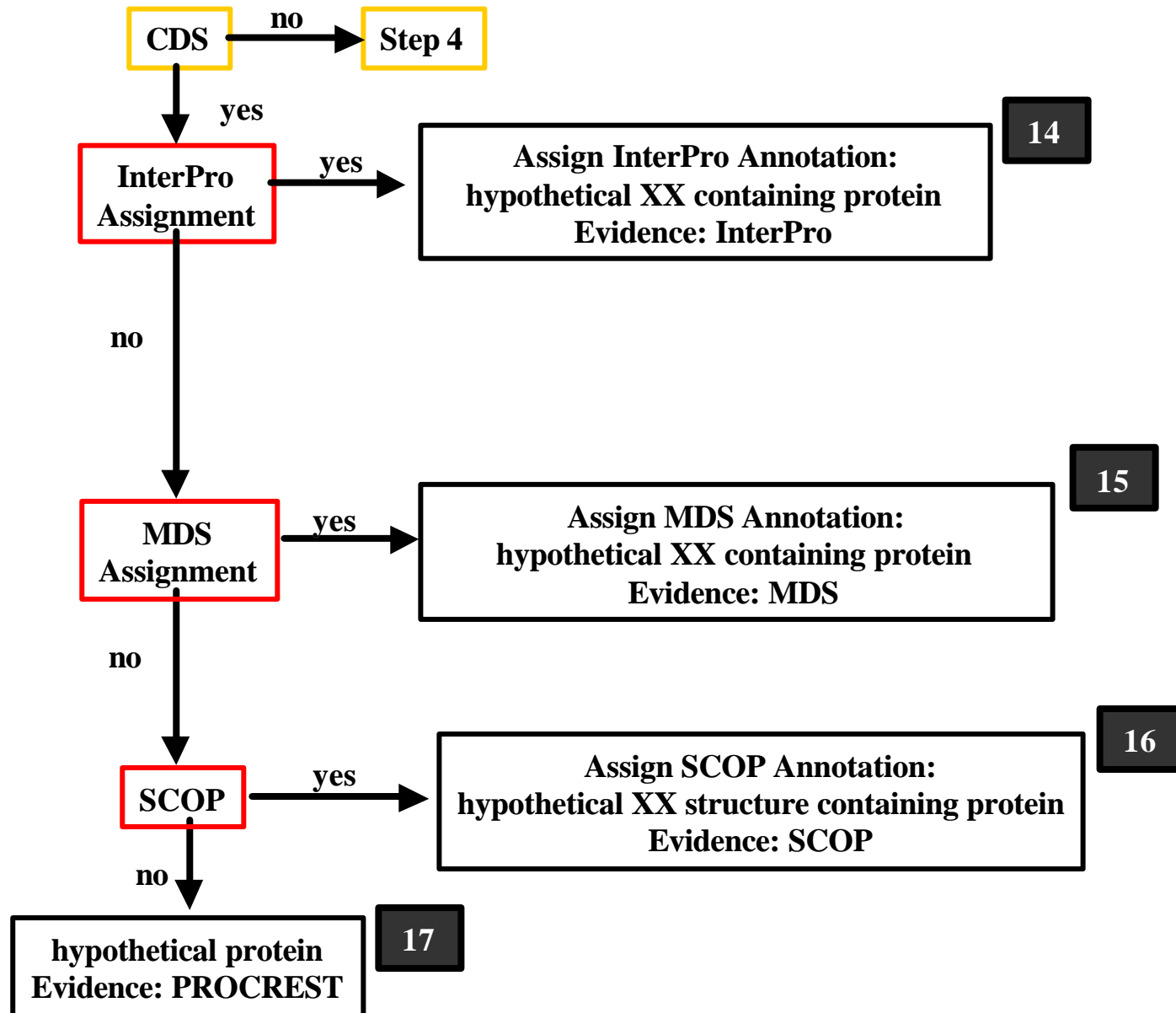
<u>Step 4: Inferred Homology-based Assignments</u>

| UniGene Assignment | →yes | TIGR_MGI Assignment | →yes | Inferred: Cluster Name Evidence: TGI/UG **11** |

UniGene Assignment —no→ (down)

TIGR_MGI Assignment —no→ Inferred: Cluster Name Evidence: UG **12**

TIGR_MGI Assignment —yes→ Inferred: Cluster Name Evidence: TGI **13**

TIGR_MGI Assignment —no→ Step 5

<u>Step 5: EST Search</u>

BLASTN vs. Mouse ESTs —yes→ Match >95% ID, >100 bp length —yes→ Unknown EST Evidence: BLASTN **18**

BLASTN vs. Mouse ESTs —no→ UnClassifiable **19**

Supplementary Information 6F
Gene Name Nomenclature

1. Gene Name: retrieve MGI gene name / gene symbol
   DB Ref: DB|Accession
   GO Assignment: (retrieve from top protein hit SP/InterPro and MGI excluding J:60000)

2. Gene Name: Do text query using Locus Link and transfer name/Gene Symbol
   Qualifier: complete
   Evidence: BLASTN, %ID, match length
   DB Ref: DB|Accession
   GO Assignment: (retrieve from top protein hit SP/InterPro and MGI excluding J:60000)

3. Gene Name: Do text query using Locus Link and transfer name/Gene Symbol
   Qualifier: partial/splice variant/unspliced/problem
   Evidence: BLASTN, %ID, match length
   DB Ref: DB|Accession
   GO Assignment: (retrieve from top protein hit SP/InterPro and MGI excluding J:60000)

4. Gene Name: <Assign protein name from def line of hit>
   Qualifier: complete
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

5. Gene Name: <Assign protein name from def line of hit>, homolog [species]
   Qualifier: complete
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

6. Gene Name: <Assign protein name from def line of hit>, homolog [species]
   Qualifier: partial/splice variant/unspliced/problem (need radio buttons on page)
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

7. Gene Name: Assign "similar to" <Assign protein name from def line of hit>, [species]
   Qualifier:complete
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

8. Gene Name: Assign "similar to" <Assign protein name from def line of hit>, [species]
   Qualifier: partial/splice variant/unspliced/problem (need radio buttons on page)
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

9. Gene Name: Assign "weakly similar to" <Assign protein name from def line of hit>, [species]
   Qualifier:complete
   Evidence: FASTY, %ID, %length, match length
   DB Ref: DB|Accession
   GO Assignment: retrieve from SP/Interpro

10. **Gene Name: Assign "weakly similar to" <Assign protein name from def line of hit>, [species]**
    Qualifier: partial/splice variant/unspliced/problem (need radio buttons on page)
    Evidence: FASTY, %ID, %length, match length
    DB Ref: DB|Accession
    GO Assignment: retrieve from SP/Interpro

11. **Gene Name: "inferred:"<cluster name>**
    Qualifier:3' UTR, 5' UTR
    Evidence: UG/TGI

12. **Gene Name: "inferred:"<cluster name>**
    Qualifier: 3' UTR, 5' UTR
    Evidence: UG

13. **Gene Name: "inferred:"<cluster name>**
    Qualifier: 3' UTR, 5' UTR
    Evidence: TGI

14. **Gene Name: hypothetical <domain> containing protein**
    Qualifier:
    Evidence: InterPro
    DB Ref: InterPro|Accession
    GO Assignment: retrieve from Interpro

15. **Gene Name: hypothetical <domain> containing protein**
    Qualifier:
    Evidence: MDS
    DB Ref: MDS|Accession

16. **Gene Name: hypothetical <domain> structure containing protein**
    Qualifier:
    Evidence: SCOP
    DB Ref: SCOP|Accession
    GO Assignment: retrieve from SCOP

17. **Gene Name: hypothetical protein**
    Qualifier:
    Evidence: DECODER/ProCrest/rsCDS/NCBI CDS Predictor/Longest ORF/Truncated-Longest ORF

18. **Gene Name: unknown EST**
    Qualifier:
    Evidence: BLASTN, %ID, match length
    DB Ref: DB|Accession

19. **Gene Name: unclassifiable**
    Qualifier:
    Evidence:
    DB Ref: