



**Supporting Figure 8. Refinement of enriched GO or SNOMED terms** (related to **Figures 2B-C, 3C, Tables S3 and S7, Protocol S2**). **Equations 1, 2 and 3** describe the refinement algorithm over the statistically significant results in an enrichment study. **Equation 1** searches the nodes with the lowest adjusted  $P$ -value,  $V_{RM}=\{V_n, V_p, V_q\}$  (red nodes). **Equation 2** excludes parent nodes that have the same adjusted  $P$ -value as their child and retains the child nodes in  $V_{RM}$ ,  $V_{RRM}=\{V_p, V_q\}$  (yellow nodes). **Equation 3** identifies the subsumed significant associations of  $V_{RRM}$ ,  $V_{SDRRM}=\{V_s, V_t, V_u, V_v, V_w\}$  (blue nodes). **Equation 4** is  $V_{included}=\{V_p, V_q, V_s, V_t, V_u, V_v, V_w\}$ , the union of  $V_{RRM}$  and  $V_{SDRRM}$  as the final subset of retained significant results.

**Legend:** Entities of an ontology (GO terms or SNOMED concepts) are illustrated as circular “nodes” with an associated adjusted  $P$ -value from the gene enrichment study (e.g.  $10^{-3}$ ) and a symbol (e.g. letter “k”). Hierarchical relationships between nodes are illustrated as an arrow pointing from a parent node towards its child node. Nodes identified by an equation are highlighted with color and located inside a dotted ellipse.