**iScience, Volume *24***

## Supplemental information

# caAtlas: An immunopeptidome atlas of human cancer

**Xinpei Yi, Yuxing Liao, Bo Wen, Kai Li, Yongchao Dou, Sara R. Savage, and Bing Zhang**

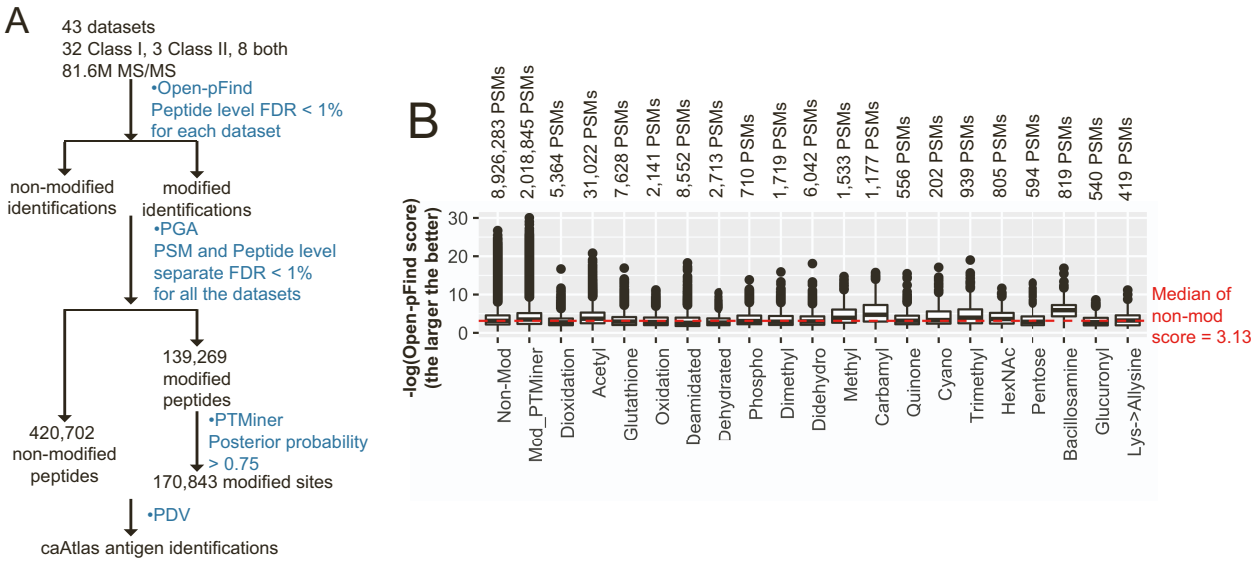


**Figure S1. Quality control of caAtlas antigen identification, related to Figure 2 and Figure 3.**

(A) Quality control pipeline for caAtlas antigen identification. (B) Score distributions of non-modified identifications (Non-Mod), modified identifications (Mod_PTMiner), and the top 19 PTM identifications shown in Figure 3C.

**Figure S2: Representative PSMs for the 19 most frequently identified PTM types, related to Figure 3.** (A) Dioxidation; (B) Acetyl; (C) Glutathione; (D) Oxidation; (E) Deamidated; (F) Dehydrated; (G) Phospho; (H) Dimethyl; (I) Didehydro; (J) Methyl; (K) Carbamyl; (L) Quinone; (M) Cyano; (N) Trimethyl; (O) HexNAc; (P) Pentose; (Q) Bacillosamine; (R) Glucuronyl; (S) Lys->Allysine.

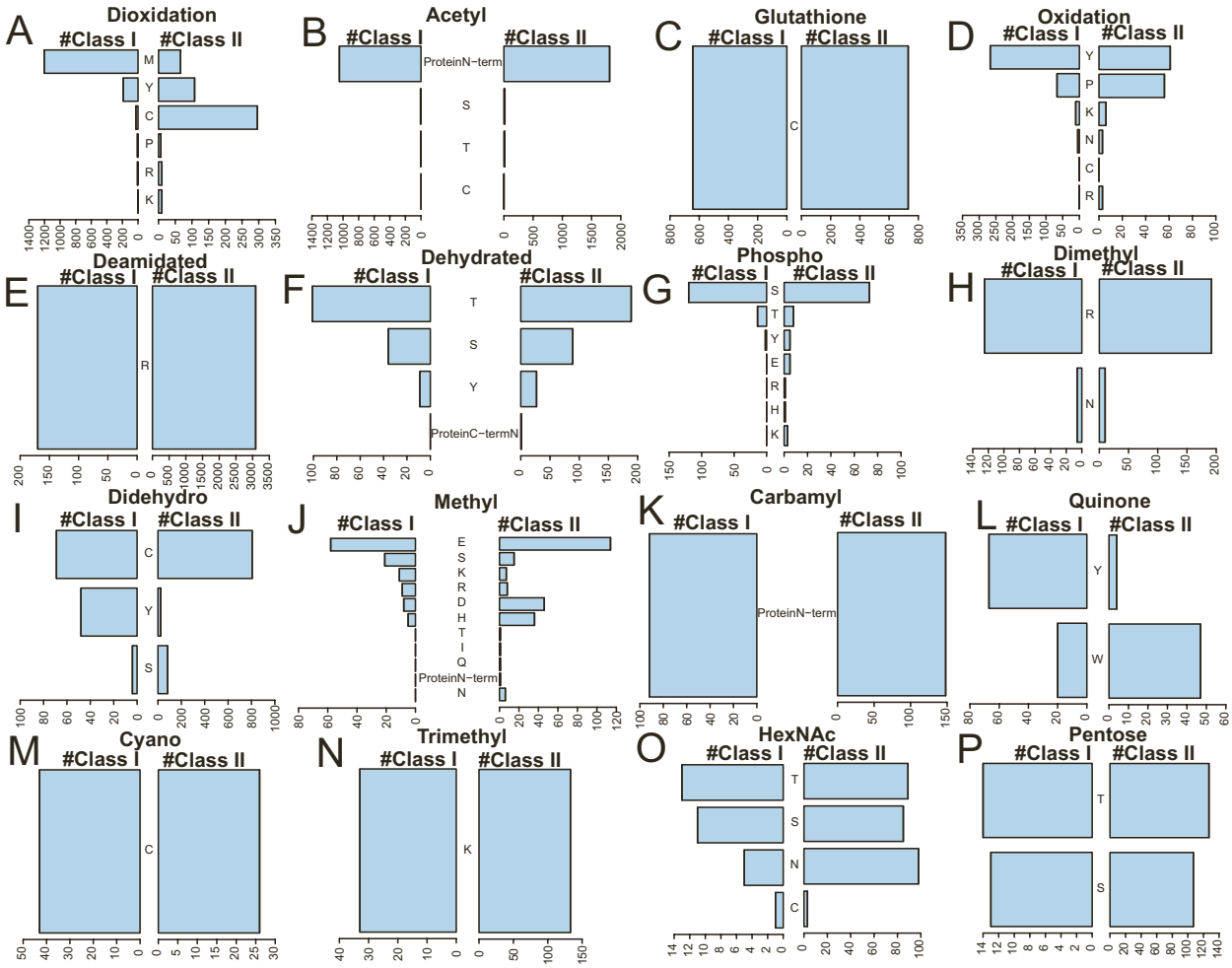


**Figure S3. Amino acid preferences of the PTM sites in HLA class I and class II PTM antigens for the 18 most frequently identified PTM types, related to Figure 4.** (A) Dioxidation; (B) Acetyl; (C) Glutathione; (D) Oxidation; (E) Deamidated; (F) Dehydrated; (G) Phospho; (H) Dimethyl; (I) Didehydro; (J) Methyl; (K) Carbamyl; (L) Quinone; (M) Cyano; (N) Trimethyl; (O) HexNAc; (P) Pentose.

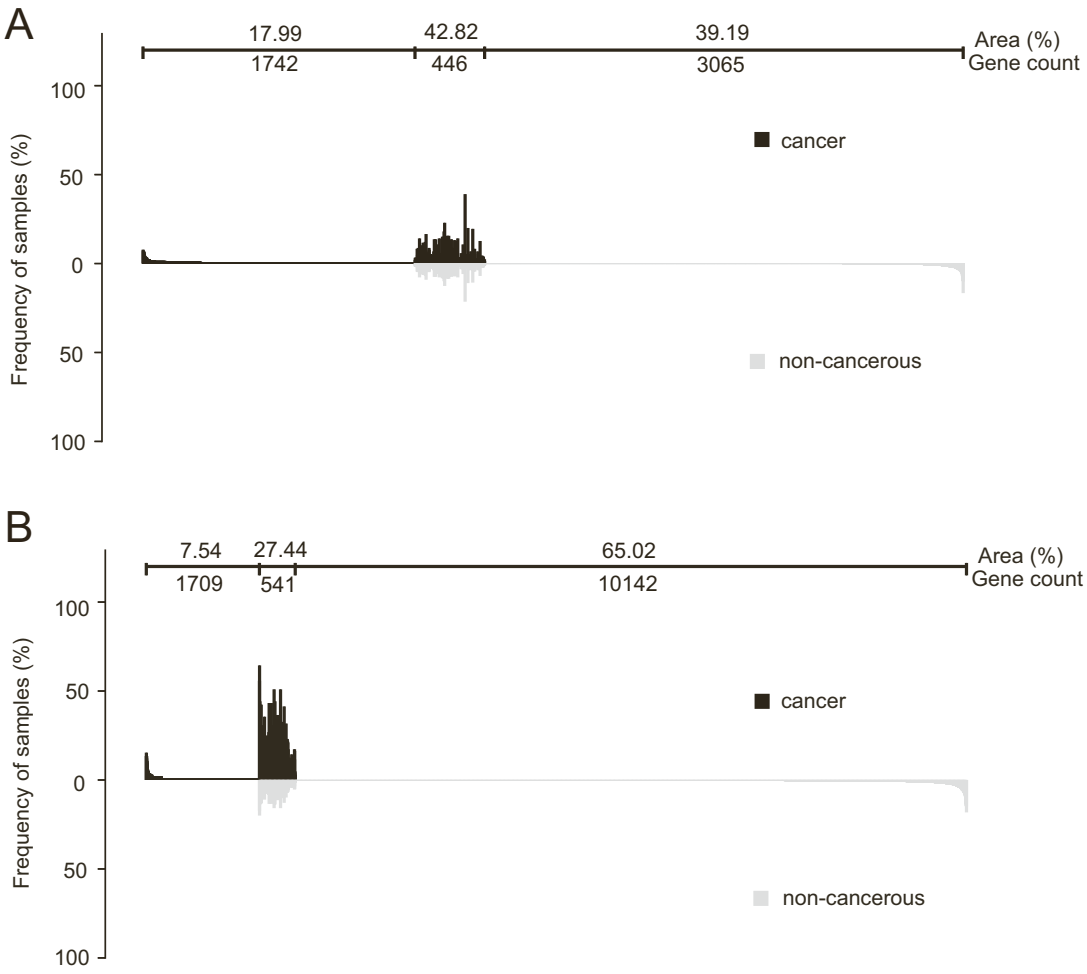


**Figure S4. Differential immunopeptidome analysis between cancer and non-cancerous samples for the 5,253 HLA class I and 12,392 HLA class II PTM antigens identified in our study, related to Figure 5.** Comparison is based on the frequency of samples for one specific PTM antigen, as indicated on the y axis. The numbers of PTM antigens that are unique to tumor samples (1,742 for HLA class I and 1,709 for HLA class II), unique to non-cancerous samples (3,065 for HLA class I and 10,142 for HLA class II) and shared between tumor and non-cancerous samples (446 for HLA class I and 541 for HLA class II) are indicated at the top of the figures, together with the respective area under the curve in percentage of total area (i.e., the sum of sample frequency for each PTM antigen divided by that for all PTM antigens). PTM antigens are sorted according to the sample numbers within each of the three groups.

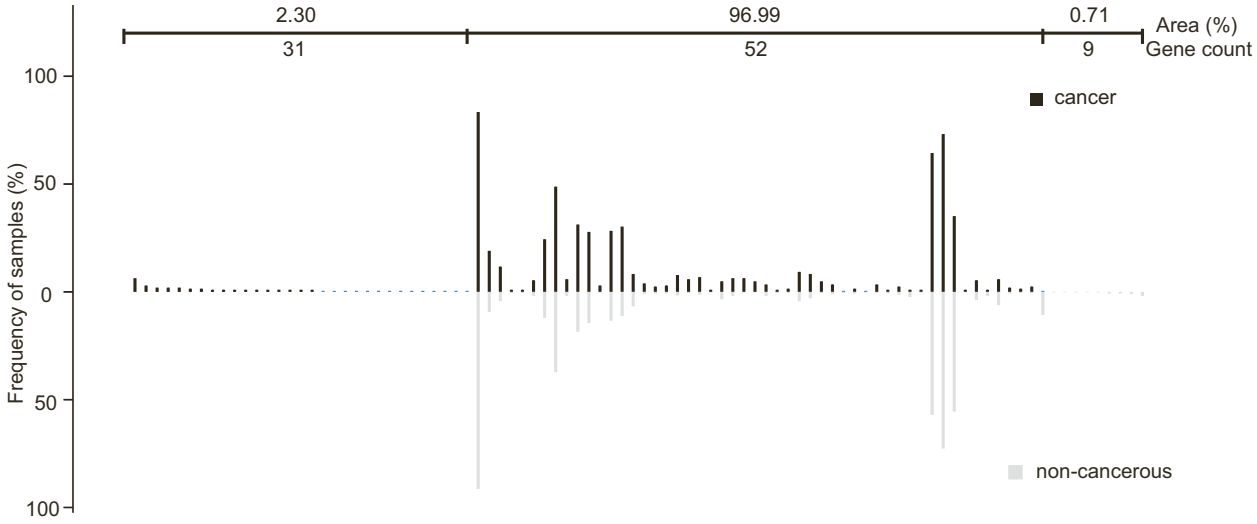


**Figure S5. Differential immunopeptidome analysis between cancer and non-cancerous samples for the 92 putative CT antigen source genes with evidence of MHC-bound peptides in our analysis, related to Figure 6.** Comparison is based on the frequency of samples for one specific putative CT antigen, as indicated on the y axis. The numbers of putative CT antigens that are unique to tumor samples (31), unique to non-cancerous samples (9), and shared between tumor and non-cancerous samples (52) are indicated at the top of the graph, together with the respective area under the curve in percentage of total area (i.e., the sum of sample frequency for each putative CT antigen divided by that for all putative CT antigens). Putative CT antigens are sorted according to the sample numbers within each of the three groups.

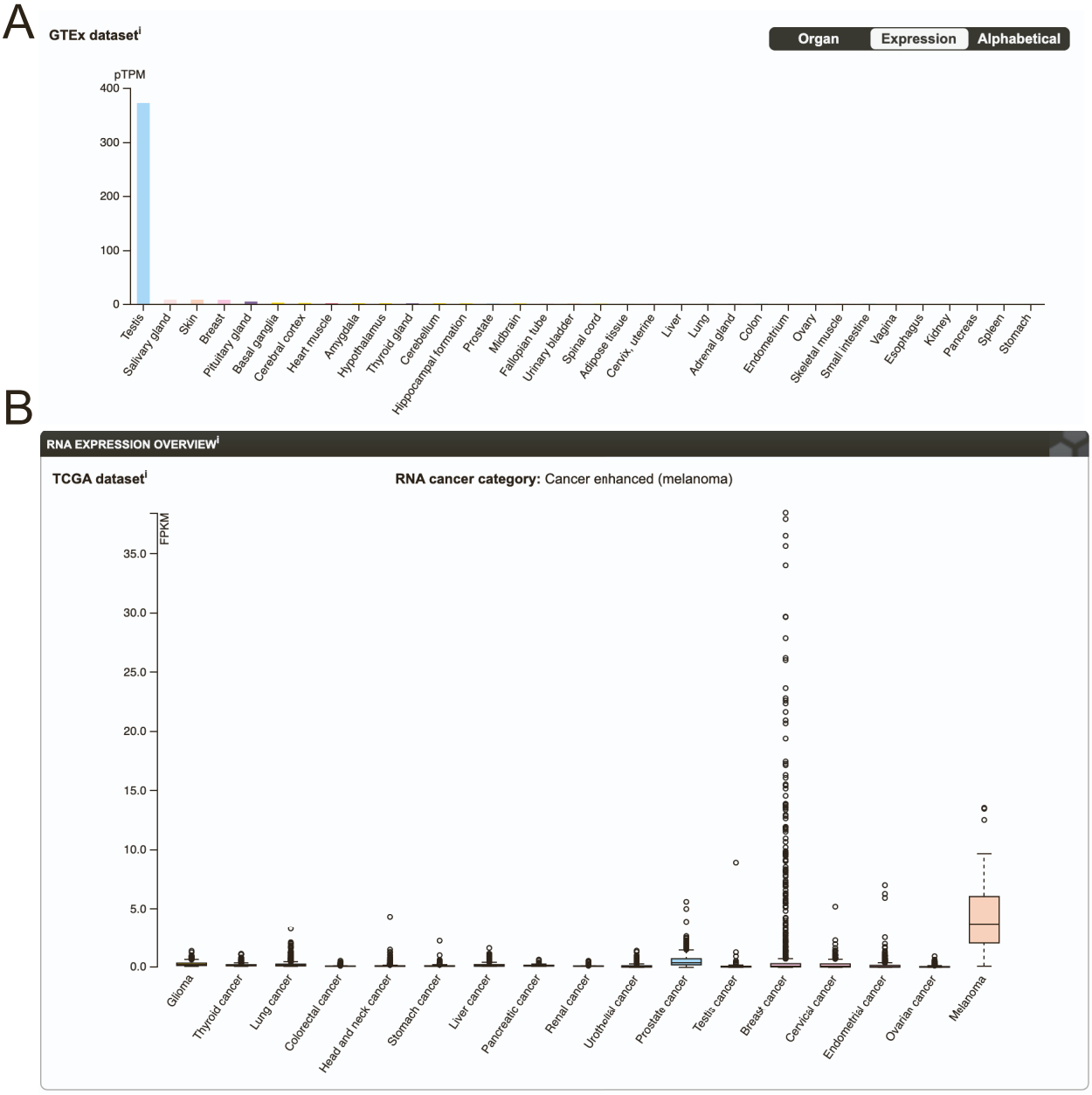


**Figure S6. ROPN1B shows restricted mRNA expression in testis among all normal tissues and is selectively expressed in melanoma among different cancer types, related to Figure 7.** (A) mRNA expression of ROPN1B in different normal tissues from the GTEx dataset downloaded from the human protein atlas website. (B) mRNA expression of ROPN1B in different cancer types from the TCGA dataset downloaded from the human protein atlas website.

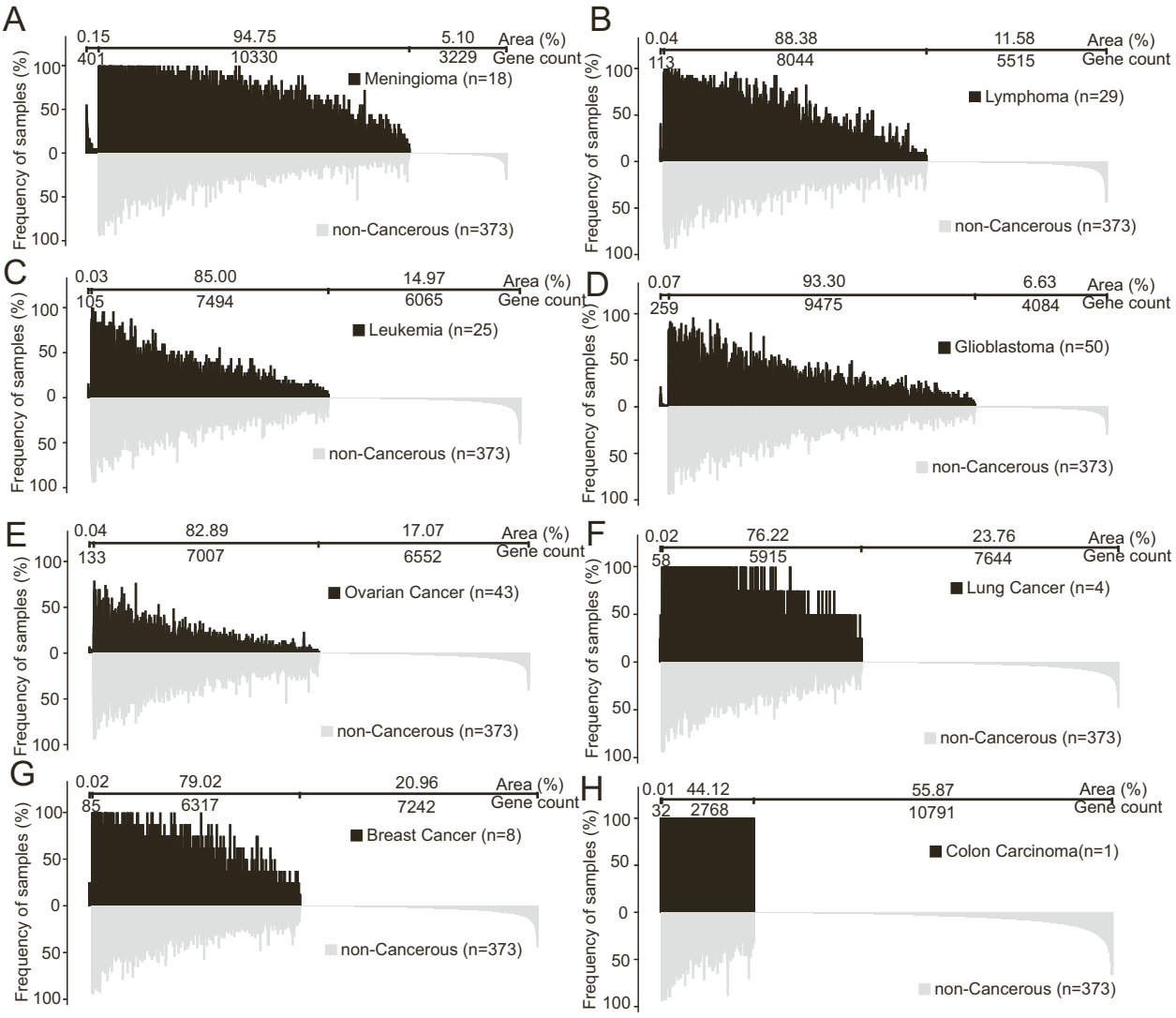


**Figure S7: Differential analysis of the antigen source gene detection frequencies between samples of a specific cancer type and all non-cancerous samples, related to Figure 7.** (A-H) Plots for meningioma, lymphoma, leukemia, glioblastoma, ovarian cancer, lung cancer, breast cancer, and colon cancer, respectively. These plots are equivalent to Figure 7A for melanoma.


**Figure S8: Neoantigen identification pipeline and results, related to STAR Methods.** (A) Schematic overview of the NeoQuery workflow for neoantigen identification. (B) Identification of previously reported neoantigens by NeoQuery. Spectrum ID (top row), HLA allele, predicted binding affinity, peptide-spectrum matching score, and false discovery rate are listed for each annotated MS/MS spectrum.