ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Identification of novel tumor antigens and immune subtypes in breast cancer patients for mRNA vaccine development

Supplementary Files

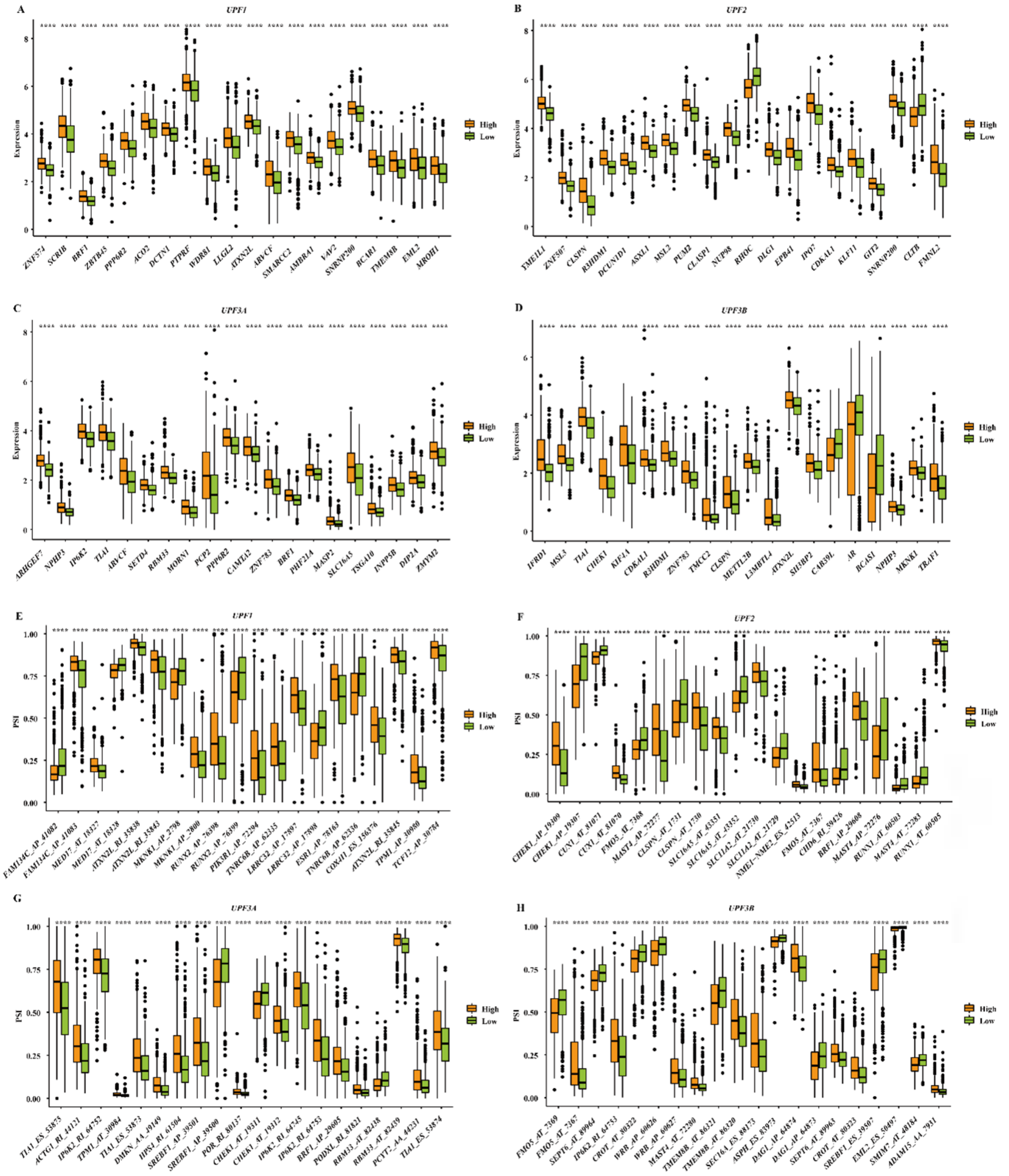

Figure S1. Relationship between nonsense-mediated mRNA decay (NMD) factors and potential antigen candidate genes. (A–D) The top 20 most differentially expressed antigen candidate genes were sequenced based on the *p*-values of the *t*-tests: (A) *UPFI*; (B) *UPF2*; (C) *UPF3A*; (D) *UPF3B*. (E–H) The top 20 variable splicing events with the most significant difference distribution are shown in order according to the *p*-value of the *t*-test: (E) *UPFI*; (F) *UPF2*; (G) *UPF3A*; (H) *UPF3B*.

Notes: **p*-value < 0.05; ***p*-value < 0.01; ****p*-value < 0.001; *****p*-value < 0.0001.

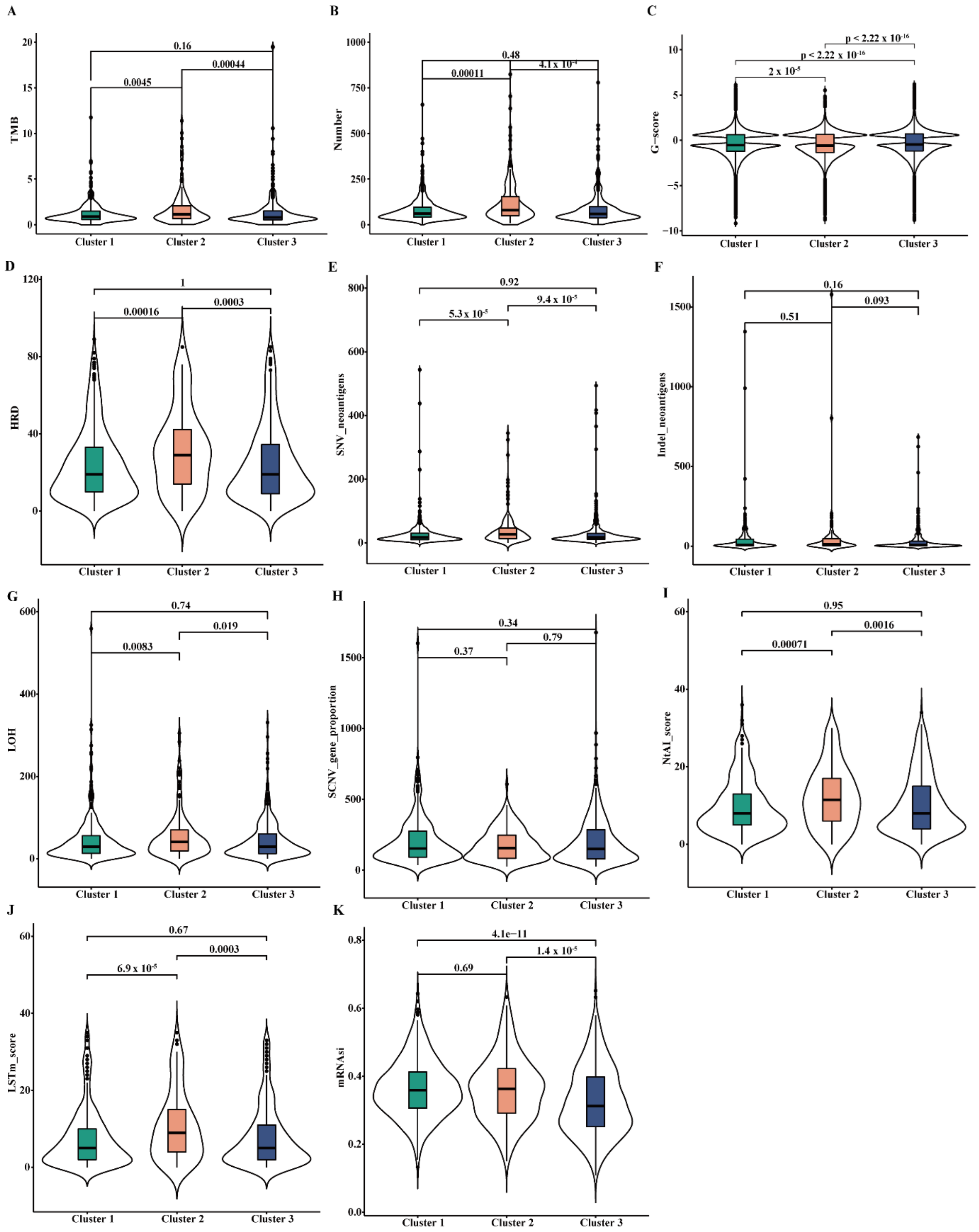

Figure S2. Mutation and copy number variation (CNV) analysis among subtypes. The Wilcox test was used to analyze the distribution of homologous recombination deficiency (HRD), neoantigen loads, chromosomal instability, and dryness index among subtypes. (A) Distribution of tumor mutation burden (TMB) scores in immune subtypes. (B) The number of genes in immune subtypes. (C) G-score distribution in subtypes. (D) HRD. (E) Single nucleotide variation (SNV), neoantigen mutation burden. (F) Insertion/deletion of a mutant load of deletion neoantigen. (G) Loss of heterozygosity (LOH) in the genome. (H) Distribution of somatic CNV (SCNV) genes. (I) Telomere allele imbalance. (J) Large fragment migration. (K) Dryness index.

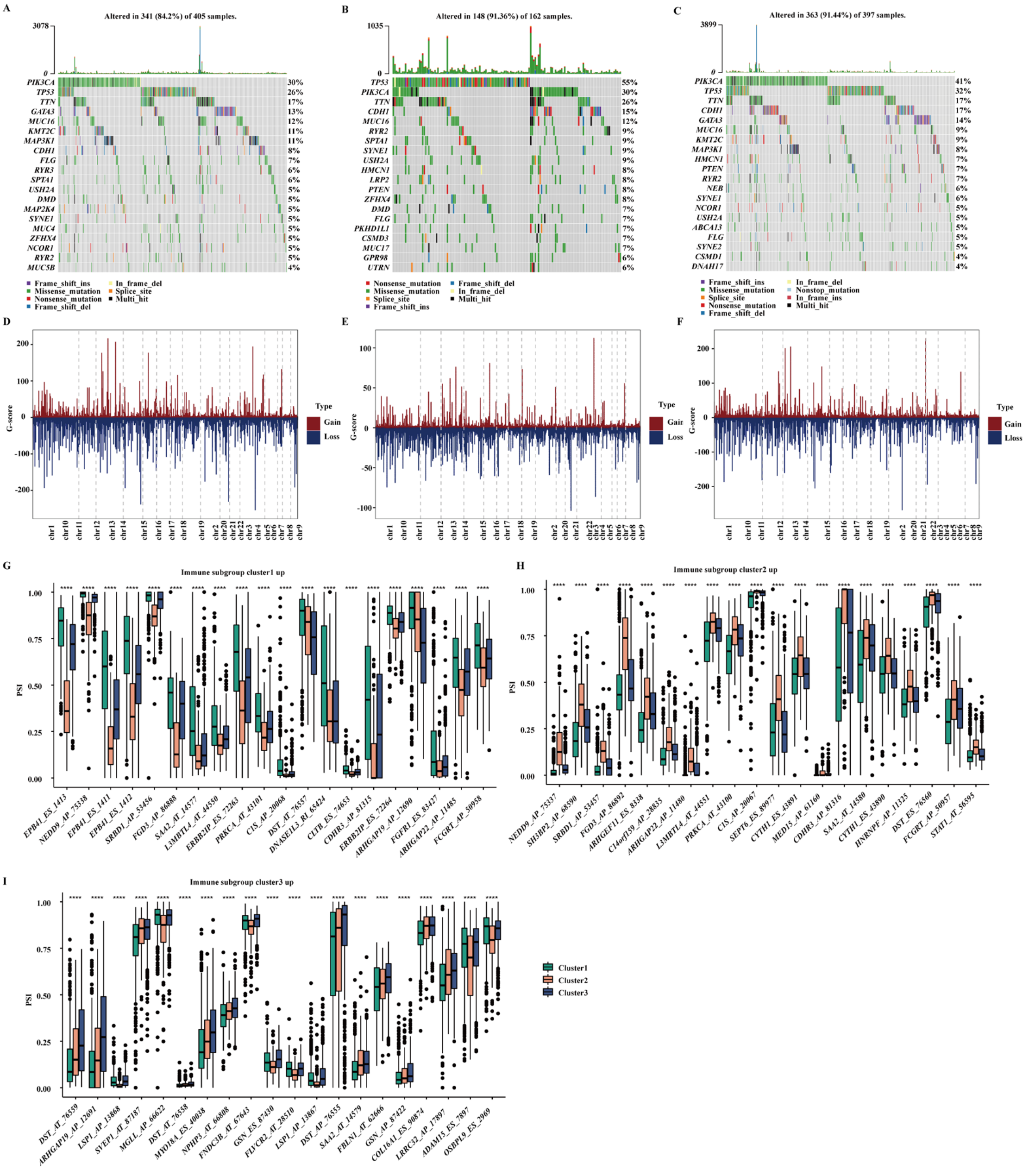

Figure S3. Alternative splicing across different patient clusters and potential antigen candidate genes. (A) Waterfall diagram showing top 20 mutations in Cluster 1 patients. (B) Waterfall diagram showing the top 20 mutations in Cluster 2 patients. (C) Waterfall diagram showing the top 20 mutations in Cluster 3 patients. (D) Chromosomal distribution of copy number in Cluster 1 patients. (E) Chromosomal distribution of copy number in Cluster 2 patients. (F) Chromosome distribution of copy number in Cluster 3 patients. (G–I) The top 20 up-regulated genes across the different subtypes were selected based on the *p*-values from the analysis of variance: (G) Cluster 1, (H) Cluster 2, (I) Cluster 3.

Notes: **p*-value < 0.05; ***p*-value < 0.01; ****p*-value < 0.001; *****p*-value < 0.0001.

Abbreviation: PSI: Percent spliced in.

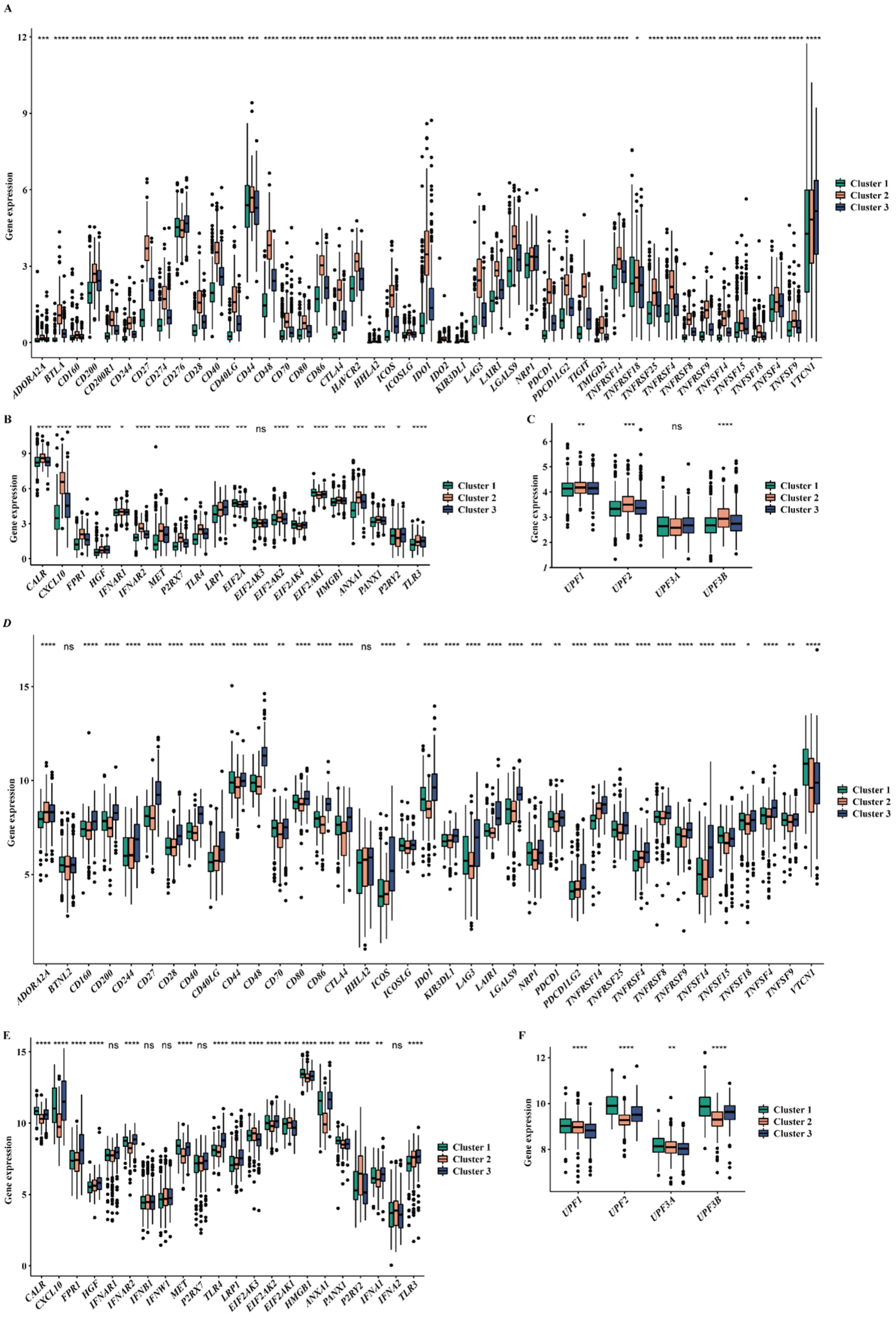

Figure S4. Expression of immune checkpoint (ICPs), immunogenic cell death (ICD), nonsense-mediated mRNA decay (NMD) factors, and immunosubtype biomarkers among subclusters of breast cancer patients. The Cancer Genome Atlas dataset: (A) ICP expression across different patient subtypes. (B) ICD expression across different patient subtypes. (C) Expression of NMD factors across different patient subtypes. (D) Expression of cancer biomarkers in different subtypes of patients. (E) Expression of the cancer gene across different patient subtypes. Validation set: (F) Expression of ICP in different subtypes of patients. (G) ICD expression across different patient subtypes. (H) Expression of NMD factor across different patient subtypes. (I) Expression of cancer biomarkers in different subtypes of patients. (J) Expression of cancer genes in different subtypes of patients.

Notes: ns: Not significant; *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001; ****p-value < 0.0001.