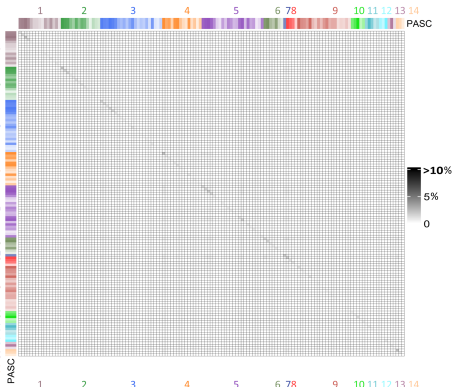
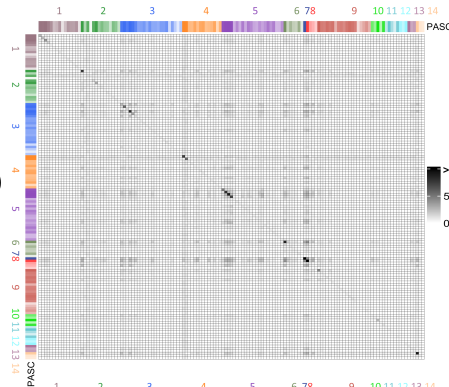

Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes

In the format provided by the
authors and unedited

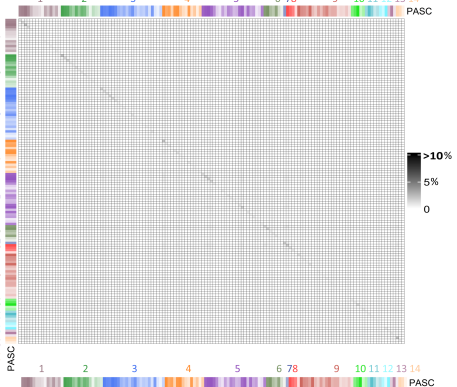
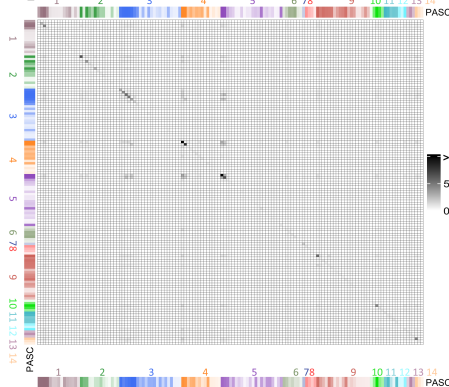
Table-of-contents

Supplementary Figure 1.....	1
Supplementary Figure 2.....	2
Supplementary Figure 3.....	3

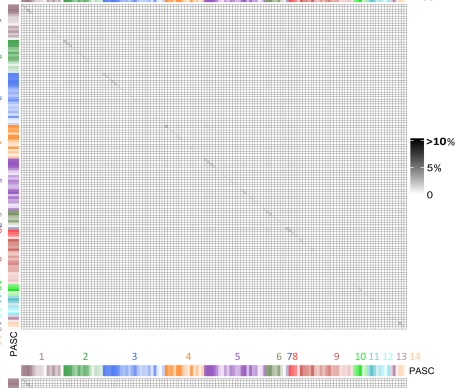
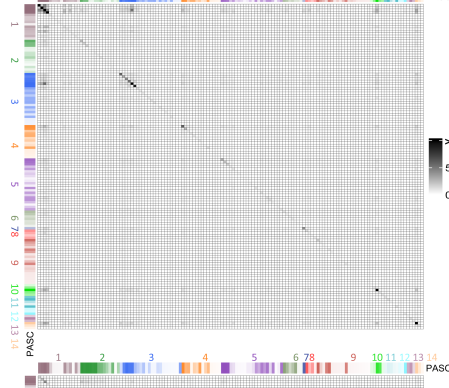
Subphenotype 1
(Cardiac and Renal)
N=7,047



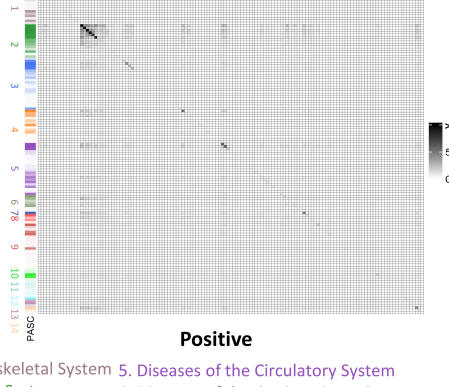
Subphenotype 2
(Respiratory, Sleep
and Anxiety)
N=6,838



Subphenotype 3
(Musculoskeletal
and Nervous)
N=4,879



Subphenotype 4
(Digestive and
Respiratory)
N=2,117



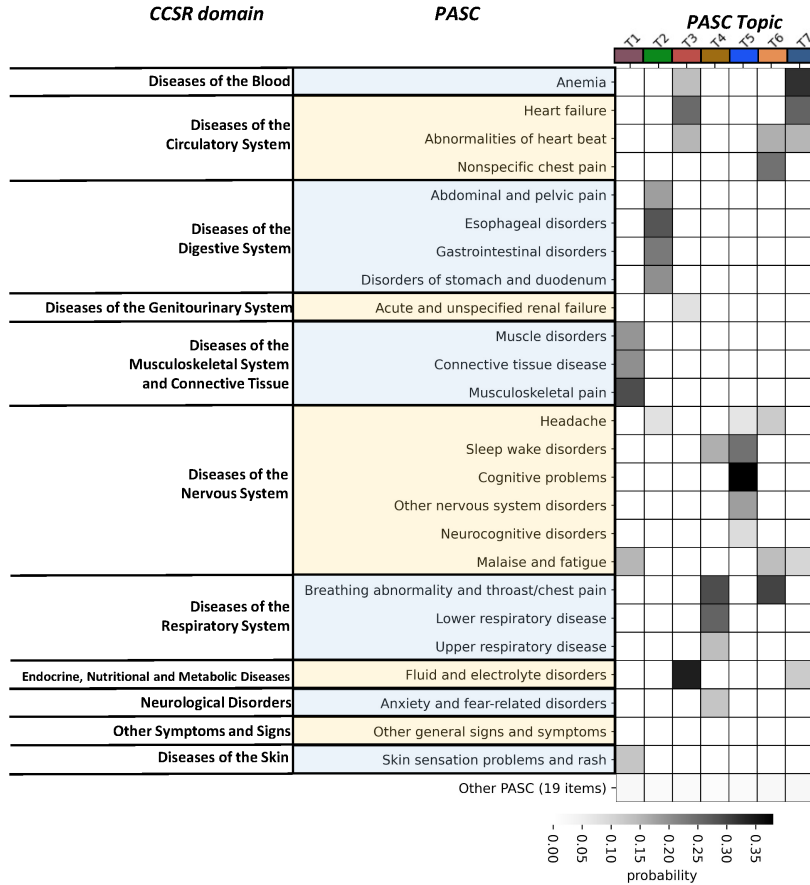
Positive

Negative

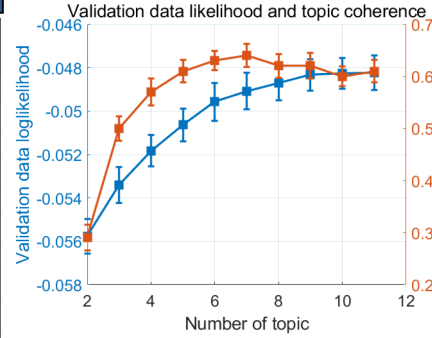
1: Diseases of the Musculoskeletal System 5: Diseases of the Circulatory System 9: Mental, Behavioral Disorders 13: Certain Infectious and Parasitic Diseases
2: Diseases of the Digestive System 6: Diseases of the Genitourinary System 10: Diseases of the Skin and Subcutaneous Tissue 14: General and others
3: Diseases of the Nervous System 7: Diseases of the Blood 11: Diseases of the Ear and Mastoid Process
4: Diseases of the Respiratory System 8: Endocrine, Nutritional and Metabolic Diseases 12: Injury, Poisoning

Supplementary Figure 1. Difference of the incidence patterns of all PASC conditions on the INSIGHT cohort (extension of Figure 4). The order of PASC follows the one provided in the page “137 PASC list” in the Supplementary Table 2.

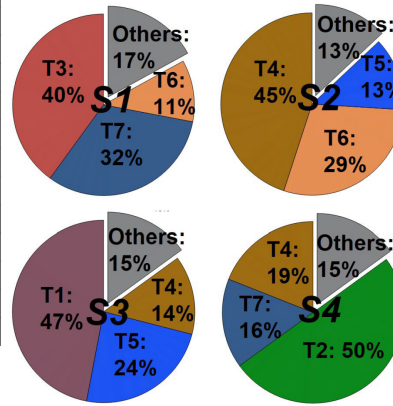
a. PASC topics learned based on the selective 44 PASC



b. Select the optimal number of topics



c. Distribution of topics on each subphenotype



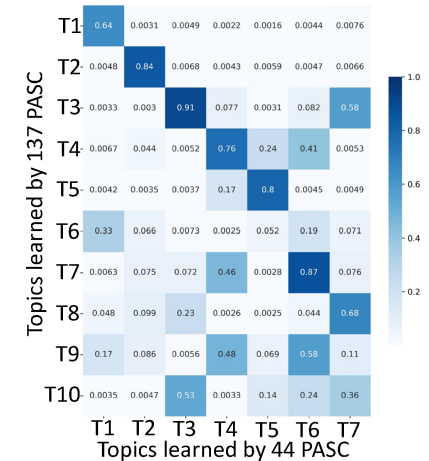
d. Overlap matrix of the subphenotypes (number of patients)

Subphenotypes by 44 PASC

	S1	S2	S3	S4
S1	6493	267	162	125
S2	326	6170	249	93
S3	104	206	4430	139
S4	24	60	44	1989

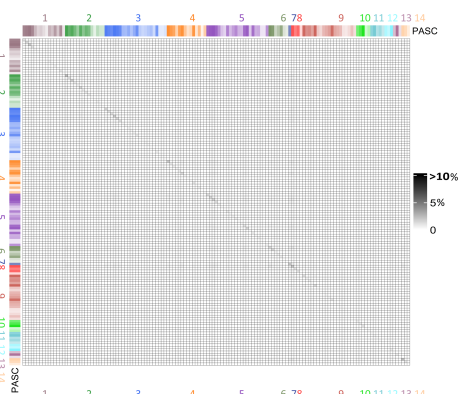
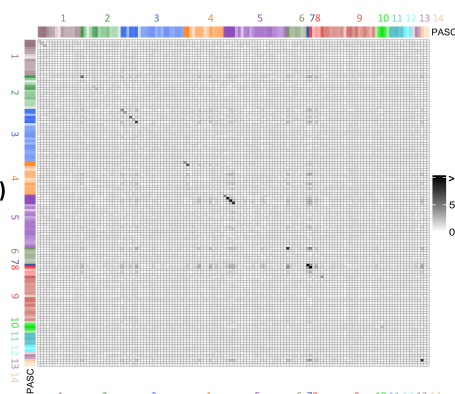
Subphenotypes by 137 PASC

e. Similarities of topic learned from 137 and 44 PASC

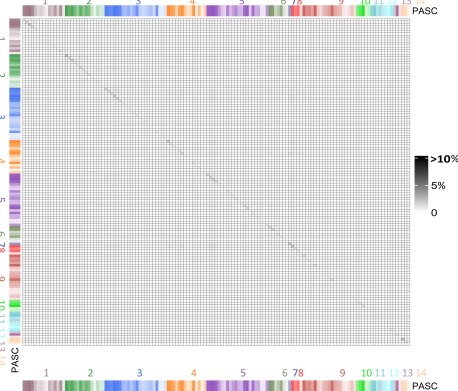
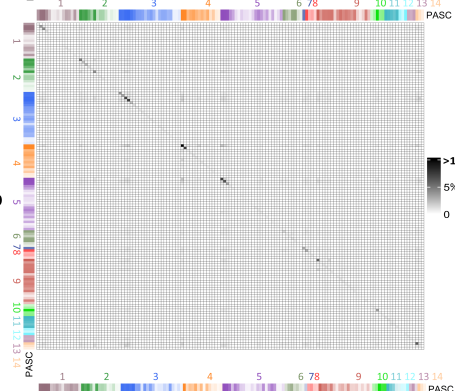


Supplementary Figure 2. We learned topics and derived subphenotypes based on the 44 selective PASC in the INSIGHT cohort. (a) shows seven topics learned based on the 44 selective PASC. (b) shows the procedure of selection optimal number of topics, where the error bar is obtained based on 5-fold cross-validation ($n_{\text{training}}=16705$, $n_{\text{validation}}=4176$), and data are presented as mean values \pm SD. (c) shows the distributions of mean topic proportions on each subphenotype. (c) shows the mean topic proportions on each subphenotype; (d) shows the overlap matrix of subphenotypes, where the number in i -th row j -th column denotes the number of patients who belongs to the i -th subphenotype derived by 137 PASC but belongs to the j -th subphenotype derived by 44 PASC. (d) shows the similarities of topics learned from 137 PASC and 44 PASC.

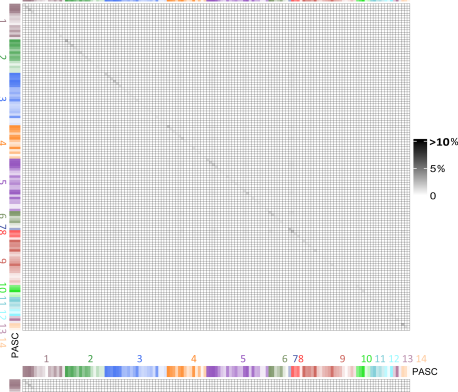
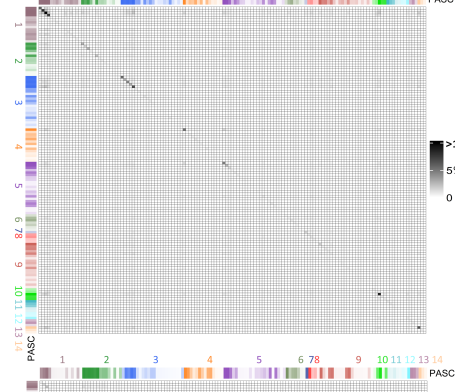
**Subphenotype 1
(Cardiac and Renal)
N=3,490**



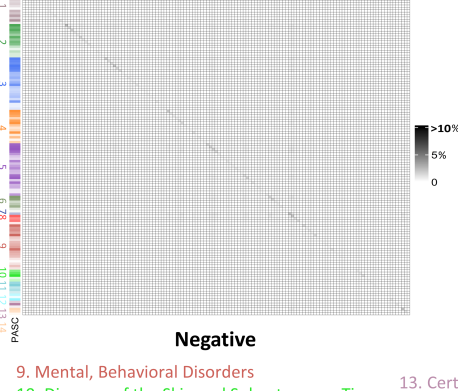
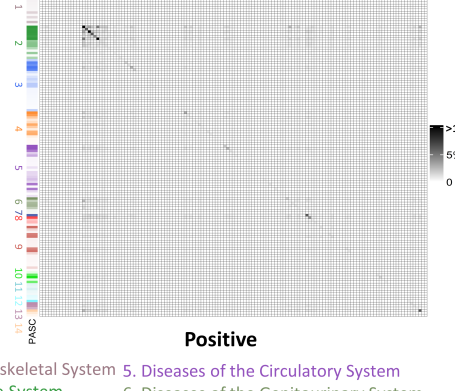
**Subphenotype 2
(Respiratory, Sleep
and Anxiety)
N=5,281**



**Subphenotype 3
(Musculoskeletal
and Nervous)
N=3,205**



**Subphenotype 4
(Digestive and
Respiratory)
N=1,748**



Positive

Negative

- | | | | |
|---|--|--|---|
| 1: Diseases of the Musculoskeletal System | 5. Diseases of the Circulatory System | 9. Mental, Behavioral Disorders | 13. Certain Infectious and Parasitic Diseases |
| 2: Diseases of the Digestive System | 6. Diseases of the Genitourinary System | 10. Diseases of the Skin and Subcutaneous Tissue | 14. General and others |
| 3: Diseases of the Nervous System | 7. Diseases of the Blood | 11. Diseases of the Ear and Mastoid Process | |
| 4. Diseases of the Respiratory System | 8. Endocrine, Nutritional and Metabolic Diseases | 12. Injury, Poisoning | |

Supplementary Figure 3. Difference of the incidence patterns of all PASC conditions on the OneFlorida+ cohort (extension of Extended Data Figure 8). The order of PASC follows the one provided in the page “137 PASC list” in the Supplementary Table 2.