

Supporting Information

**A peptide resource for the analysis of *Staphylococcus aureus* in
host–pathogen interaction studies**

c

Maren Depke, Stephan Michalik, Alexander Rabe, Kristin Surmann,
Lars Brinkmann, Nico Jehmlich, Jörg Bernhardt, Michael Hecker,
Bernd Wollscheid, Zhi Sun, Robert L. Moritz, Uwe Völker and Frank Schmidt

Supplemental Material S1: Details of LC-MS/MS analysis and database search

strong cationic exchange chromatography (SCX)	
instrument	Ultimate 3000 RSLC (Dionex, Idstein, Germany)
SCX column	nano column (Poros 10S, 300 µm x 10 cm, nanoViper, SCX, 10 µm)
buffer system	binary buffer system consisting of 5 mM Na ₂ HPO ₄ pH 3, 5% ACN in water (SCX buffer A) and 1 M NaCl, 5 mM Na ₂ HPO ₄ pH 3, 5% ACN in water (SCX buffer B)
flow rate	loading: 5 µl/min elution: 300 nl/min
method	2D salt steps; 7 different concentrations of NaCl
salt step volume	12 µl
salt step concentrations	2 mM, 10 mM, 20 mM, 50 mM, 100 mM, 200 mM, 500 mM
column oven temperature	40°C

reversed phase liquid chromatography (RPLC)	
instrument	Ultimate 3000 RSLC (Dionex, Idstein, Germany)
trap column	75 µm inner diameter, packed with 3 µm C18 particles (Acclaim PepMap100, Thermo Scientific)
analytical column	<i>for SCX and off-gel pre-fractionated samples and non-fractionated samples (standard method):</i> 15 cm analytical column packed with 2 µm C18 particles (Acclaim PepMap RSLC, Thermo Scientific) <i>for samples from cell culture infection model:</i> 25 cm analytical column packed with 2 µm C18 particles (Acclaim PepMap RSLC, Thermo Scientific)
buffer system	binary buffer system consisting of 0.1% acetic acid, 2% ACN (buffer A) and 0.1% acetic acid in 100% ACN in (buffer B)
flow rate	300 nl/min
gradient	linear gradient of buffer B from 2% up to 25%
gradient duration	30 min <i>for SCX pre-fractionated samples</i> 30 min <i>for off-gel pre-fractionated samples</i> 100 min <i>for non-fractionated samples (standard method)</i> 120 min <i>for samples from cell culture infection model</i>
column oven temperature	40°C

mass spectrometry (MS)

instrument	Q Exactive mass spectrometer (Thermo Scientific)
electrospray	via TriVersa NanoMate (Advion Biosciences, Norwich, UK)
operation mode	data-dependent
MS scan resolution	70,000
MS ion target value	3e6
maximum ion injection time for the MS scan	250 ms <i>for SCX pre-fractionated, off-gel pre-fractionated, and non-fractionated samples (standard method)</i> 120 ms <i>for samples from cell culture infection model</i>
selection for MS/MS	10 most abundant isotope patterns with charge ≥ 2 from the survey scan
isolation window	3 <i>m/z</i>
dissociation mode	higher energy collisional dissociation (HCD)
normalized collision energy	27.5%
maximum ion injection time for the MS/MS scans	120 ms
MS/MS ion target value	2e5
dynamic exclusion	40 s <i>for SCX pre-fractionated, off-gel pre-fractionated, and non-fractionated samples (standard method)</i> 30 s <i>for samples from cell culture infection model</i>
MS/MS operation mode	profile mode, peptide match: on <i>for SCX pre-fractionated, off-gel pre-fractionated, and non-fractionated samples (standard method)</i> centroid mode, peptide match: preferred <i>for samples from cell culture infection model</i>

database search and analysis

SEARCH PARAMETERS AND ACCEPTANCE CRITERIA

<i>search engine</i> name version or release date	MASCOT Server / MASCOT Daemon version 2.3.2
<i>sequence database</i> name version release date number of entries	Sau_8325_BLAST_HGW S.Aureus_DB_NCTC8325_NC007795_BLAST_HGW_20130327.fasta 2013-03-27 2891 database without reversed entries
<i>spectral library</i> name version source of sequences source of spectra software used for library generation	20130930_StaphMascot_noProb 2013-09-30 FASTA-file, see sequence database above 144 raw files of this study Trans Proteomics Pipeline (TPP)
<i>enzyme specificity</i> enzyme missed cleavages non-specific cleavages	trypsin max. 2 allowed none
<i>modifications</i> fixed modifications variable modifications	none deamidation carbamidomethylation (C) oxidation (M) conversion of Gln to pyroGlu (N-term Q)
<i>mass tolerance</i> for precursor ions for fragment ions	10 ppm 0.1 Da
known contaminants excluded	not excluded
<i>threshold score</i> used for accepting individual spectra	<p>data from MASCOT search: ion score > 20 used for all analyzed files</p> <p>For each file, an ion score >20 is higher than the threshold score which indicates identity or extensive homology with $p < 0.05$. Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999, 20, 3551-3567.</p> <p>Individual ion scores at the threshold of $p < 0.05$ for the individual files are indicated below; minimum = 13; maximum = 18; mean = 16.5; median = 17; standard deviation = 0.97 (calculated from all data sets).</p> <p>data from spectral library search: DOT score > 0.5 and absolute precursor tolerance < 0.01</p>

threshold score
used for accepting individual spectra
(continued)

file	ion score threshold with $p < 0.05$
BR2_F1	17
BR2_F2	17
BR2_F3	18
BR2_F4	17
BR2_F5	17
BR2_F6	17
BR2_F7	16
BR2_F8	16
BR2_F9	17
BR2_F10	17
BR2_F11	17
BR2_F12	17
BR2_F13	16
BR2_F14	16
BR2_F15	16
BR2_F16	16
BR2_F17	16
BR2_F18	15
BR2_F19	15
BR2_F20	16
BR2_F21	16
BR2_F22	16
BR2_F23	16
BR2_F24	15
BR3_1	16
BR3_2	17
BR3_3	17
BR3_4	17
BR3_5	17
BR3_6	16
BR3_7	15
BR3_8	15
BR3_9	16
BR3_10	16
BR3_11	16
BR3_12	16
BR3_13	15
BR3_14	15
BR3_15	15
BR3_16	15
BR3_17	14
BR3_18	13
BR3_19	13
BR3_20	14
BR3_21	15
BR3_22	15
BR3_23	15
BR3_24	14
HG001_AOFS_BR1_F1	17
HG001_AOFS_BR1_F2	17
HG001_AOFS_BR1_F3	17
HG001_AOFS_BR1_F4	17
HG001_AOFS_BR1_F5	17
HG001_AOFS_BR1_F6	17
HG001_AOFS_BR1_F7	16
HG001_AOFS_BR1_F8	17
HG001_AOFS_BR1_F9	17
HG001_AOFS_BR1_F10	17
HG001_AOFS_BR1_F11	17
HG001_AOFS_BR1_F12	17
HG001_AOFS_BR1_F13	17
HG001_AOFS_BR1_F14	17
HG001_AOFS_BR1_F15	17
HG001_AOFS_BR1_F16	17
HG001_AOFS_BR1_F17	16
HG001_AOFS_BR1_F18	16
HG001_AOFS_BR1_F19	16
HG001_AOFS_BR1_F20	16
HG001_AOFS_BR1_F21	16
HG001_AOFS_BR1_F22	15
HG001_AOFS_BR1_F23	16
HG001_AOFS_BR1_F24	16

<i>threshold score</i> used for accepting individual spectra (continued)	file	ion score threshold with $p < 0.05$
	HG001_AOFS_BR2_F1	17
	HG001_AOFS_BR2_F2	17
	HG001_AOFS_BR2_F3	17
	HG001_AOFS_BR2_F4	17
	HG001_AOFS_BR2_F5	17
	HG001_AOFS_BR2_F6	17
	HG001_AOFS_BR2_F7	17
	HG001_AOFS_BR2_F8	17
	HG001_AOFS_BR2_F9	17
	HG001_AOFS_BR2_F10	17
	HG001_AOFS_BR2_F11	17
	HG001_AOFS_BR2_F12	17
	HG001_AOFS_BR2_F13	17
	HG001_AOFS_BR2_F14	17
	HG001_AOFS_BR2_F15	17
	HG001_AOFS_BR2_F16	17
	HG001_AOFS_BR2_F17	16
	HG001_AOFS_BR2_F18	16
	HG001_AOFS_BR2_F19	16
	HG001_AOFS_BR2_F20	16
	HG001_AOFS_BR2_F21	16
	HG001_AOFS_BR2_F22	16
	HG001_AOFS_BR2_F23	16
	HG001_AOFS_BR2_F24	16
	HG001_AOFS_BR3_F1	17
	HG001_AOFS_BR3_F2	17
	HG001_AOFS_BR3_F3	17
	HG001_AOFS_BR3_F4	17
	HG001_AOFS_BR3_F5	17
	HG001_AOFS_BR3_F6	17
	HG001_AOFS_BR3_F7	17
	HG001_AOFS_BR3_F8	17
	HG001_AOFS_BR3_F9	17
	HG001_AOFS_BR3_F10	17
	HG001_AOFS_BR3_F11	17
	HG001_AOFS_BR3_F12	17
	HG001_AOFS_BR3_F13	17
	HG001_AOFS_BR3_F14	17
	HG001_AOFS_BR3_F15	17
	HG001_AOFS_BR3_F16	17
	HG001_AOFS_BR3_F17	16
	HG001_AOFS_BR3_F18	17
	HG001_AOFS_BR3_F19	16
	HG001_AOFS_BR3_F20	16
	HG001_AOFS_BR3_F21	16
	HG001_AOFS_BR3_F22	16
	HG001_AOFS_BR3_F23	16
	HG001_AOFS_BR3_F24	16
	Salt_Steps_2mM_AOFS_BR1_1	17
	Salt_Steps_2mM_AOFS_BR2_1	17
	Salt_Steps_2mM_AOFS_BR3_1	17
	Salt_Steps_10mM_AOFS_BR1_2	17
	Salt_Steps_10mM_AOFS_BR2_2	17
	Salt_Steps_10mM_AOFS_BR3_2	17
	Salt_Steps_20mM_AOFS_BR1_3	18
	Salt_Steps_20mM_AOFS_BR2_3	18
	Salt_Steps_20mM_AOFS_BR3_3	18
	Salt_Steps_50mM_AOFS_BR1_4	18
	Salt_Steps_50mM_AOFS_BR2_4	18
	Salt_Steps_50mM_AOFS_BR3_4	18
	Salt_Steps_100mM_AOFS_BR1_5	18
	Salt_Steps_100mM_AOFS_BR2_5	18
	Salt_Steps_100mM_AOFS_BR3_5	18
	Salt_Steps_200mM_AOFS_BR1_6	18
	Salt_Steps_200mM_AOFS_BR2_6	18
	Salt_Steps_200mM_AOFS_BR3_6	18
	Salt_Steps_500mM_AOFS_BR1_7	18
	Salt_Steps_500mM_AOFS_BR2_7	18
	Salt_Steps_500mM_AOFS_BR3_7	18
	Staph_UT_BR1	17
	Staph_UT_BR2	17
	Staph_UT_BR3	17

False Discovery Rates

Percolator score FDR

(Kall, L., Canterbury, J. D., Weston, J., Noble, W. S., MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature methods 2007, 4, 923-925.)

FDR was calculated by a target decoy approach (TDA) in combination with the Percolator score.

minimum = 0; maximum = 2.75; mean = 0.25;
median = 0.19; standard deviation = 0.32
(calculated from all data sets)

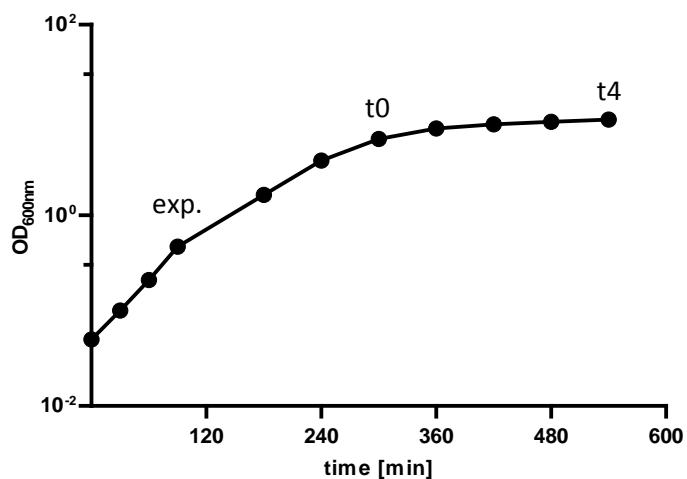
file	FDR
BR2_F1	0.31
BR2_F2	0.09
BR2_F3	0.21
BR2_F4	0.16
BR2_F5	0.3
BR2_F6	0.1
BR2_F7	0.09
BR2_F8	0.22
BR2_F9	0.22
BR2_F10	0
BR2_F11	0.24
BR2_F12	0.32
BR2_F13	0.26
BR2_F14	0.37
BR2_F15	0.65
BR2_F16	0.39
BR2_F17	0
BR2_F18	0
BR2_F19	0.32
BR2_F20	0.11
BR2_F21	0.63
BR2_F22	0.42
BR2_F23	0.42
BR2_F24	0.35
BR3_1	0.16
BR3_2	0.12
BR3_3	0.11
BR3_4	0.14
BR3_5	0.10
BR3_6	0.32
BR3_7	0.17
BR3_8	0.22
BR3_9	0.34
BR3_10	0.48
BR3_11	0.42
BR3_12	0.26
BR3_13	0.24
BR3_14	0.57
BR3_15	0.00
BR3_16	0.00
BR3_17	0.00
BR3_18	0.00
BR3_19	0.00
BR3_20	0.75
BR3_21	0.48
BR3_22	0.23
BR3_23	0.27
BR3_24	0.23
HG001_AOFS_BR1_F1	0.13
HG001_AOFS_BR1_F2	0.06
HG001_AOFS_BR1_F3	0.17
HG001_AOFS_BR1_F4	0.13
HG001_AOFS_BR1_F5	0.64
HG001_AOFS_BR1_F6	0.28
HG001_AOFS_BR1_F7	0.20
HG001_AOFS_BR1_F8	0.00
HG001_AOFS_BR1_F9	0.24
HG001_AOFS_BR1_F10	0.27
HG001_AOFS_BR1_F11	0.24
HG001_AOFS_BR1_F12	0.21
HG001_AOFS_BR1_F13	0.46
HG001_AOFS_BR1_F14	0.00
HG001_AOFS_BR1_F15	0.26
HG001_AOFS_BR1_F16	0.20
HG001_AOFS_BR1_F17	0.00
HG001_AOFS_BR1_F18	1.89
HG001_AOFS_BR1_F19	0.00
HG001_AOFS_BR1_F20	0.56
HG001_AOFS_BR1_F21	0.51

False Discovery Rates (continued)

Percolator score FDR

file	FDR
HG001_AOFS_BR1_F22	2.75
HG001_AOFS_BR1_F23	0.00
HG001_AOFS_BR1_F24	0.68
HG001_AOFS_BR2_F1	0.06
HG001_AOFS_BR2_F2	0.11
HG001_AOFS_BR2_F3	N/A
HG001_AOFS_BR2_F4	0.11
HG001_AOFS_BR2_F5	0.14
HG001_AOFS_BR2_F6	0.17
HG001_AOFS_BR2_F7	0.14
HG001_AOFS_BR2_F8	0.12
HG001_AOFS_BR2_F9	0.33
HG001_AOFS_BR2_F10	N/A
HG001_AOFS_BR2_F11	0.28
HG001_AOFS_BR2_F12	0.41
HG001_AOFS_BR2_F13	0.20
HG001_AOFS_BR2_F14	0.00
HG001_AOFS_BR2_F15	0.18
HG001_AOFS_BR2_F16	0.17
HG001_AOFS_BR2_F17	0.00
HG001_AOFS_BR2_F18	0.00
HG001_AOFS_BR2_F19	0.00
HG001_AOFS_BR2_F20	0.49
HG001_AOFS_BR2_F21	0.31
HG001_AOFS_BR2_F22	0.49
HG001_AOFS_BR2_F23	0.47
HG001_AOFS_BR2_F24	0.35
HG001_AOFS_BR3_F1	0.13
HG001_AOFS_BR3_F2	0.06
HG001_AOFS_BR3_F3	0.11
HG001_AOFS_BR3_F4	0.06
HG001_AOFS_BR3_F5	0.09
HG001_AOFS_BR3_F6	0.21
HG001_AOFS_BR3_F7	0.00
HG001_AOFS_BR3_F8	0.13
HG001_AOFS_BR3_F9	0.50
HG001_AOFS_BR3_F10	0.19
HG001_AOFS_BR3_F11	0.23
HG001_AOFS_BR3_F12	0.27
HG001_AOFS_BR3_F13	0.89
HG001_AOFS_BR3_F14	0.00
HG001_AOFS_BR3_F15	0.26
HG001_AOFS_BR3_F16	0.16
HG001_AOFS_BR3_F17	0.00
HG001_AOFS_BR3_F18	0.00
HG001_AOFS_BR3_F19	0.00
HG001_AOFS_BR3_F20	0.00
HG001_AOFS_BR3_F21	0.39
HG001_AOFS_BR3_F22	0.36
HG001_AOFS_BR3_F23	0.44
HG001_AOFS_BR3_F24	0.68
Salt_Steps_2mM_AOFS_BR1_1	N/A
Salt_Steps_10mM_AOFS_BR1_2	0.27
Salt_Steps_20mM_AOFS_BR1_3	0.25
Salt_Steps_50mM_AOFS_BR1_4	0.3
Salt_Steps_100mM_AOFS_BR1_5	0.08
Salt_Steps_200mM_AOFS_BR1_6	0.04
Salt_Steps_50mM_AOFS_BR1_7	0.11
Salt_Steps_2mM_AOFS_BR2_1	0.04
Salt_Steps_10mM_AOFS_BR2_2	0.16
Salt_Steps_20mM_AOFS_BR2_3	0.09
Salt_Steps_50mM_AOFS_BR2_4	0.16
Salt_Steps_100mM_AOFS_BR2_5	0.07
Salt_Steps_200mM_AOFS_BR2_6	0.00
Salt_Steps_50mM_AOFS_BR2_7	0.1
Salt_Steps_2mM_AOFS_BR3_1	0.47
Salt_Steps_10mM_AOFS_BR3_2	0.13
Salt_Steps_20mM_AOFS_BR3_3	0.17
Salt_Steps_50mM_AOFS_BR3_4	0.22
Salt_Steps_100mM_AOFS_BR3_5	0.12
Salt_Steps_200mM_AOFS_BR3_6	0.05
Salt_Steps_50mM_AOFS_BR3_7	0.17
Staph_UT_BR1	0.38
Staph_UT_BR2	0.48
Staph_UT_BR3	0

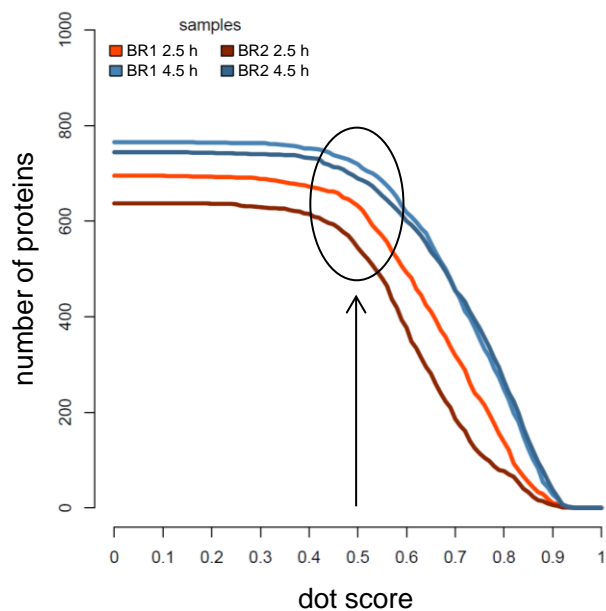
Supplemental Figure S1



Growth curve of *S. aureus* HG001 in TSB.

One representative curve of three biological replicates is depicted. Sampling points are indicated.

Supplemental Figure S2



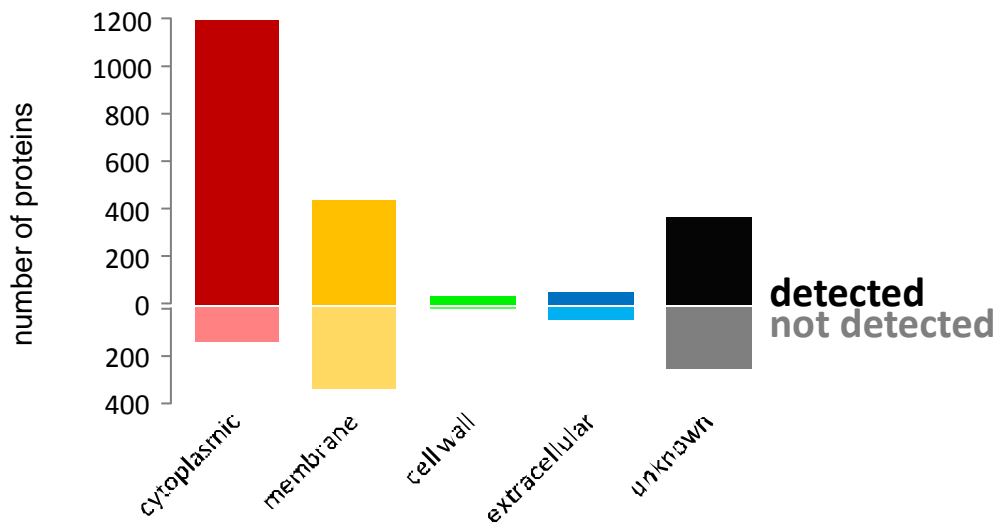
Number of identified proteins depending on score values.

The numbers of identified proteins are plotted depending on the dot score values in identification search using the spectral library generated in this study.

The arrow and circle indicate the region where the dot score starts discriminating true identifications from unspecific false identifications.

Supplemental Figure S3

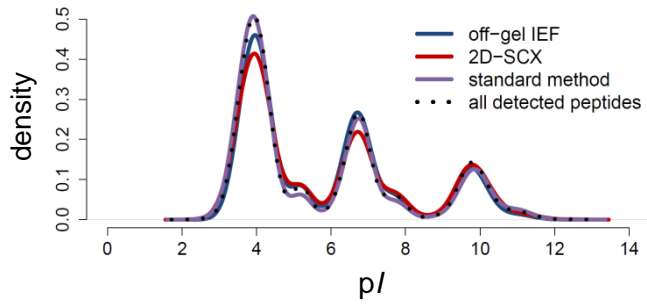
	detected proteins		proteins not detected	
	number	%	number	%
total	2088	100	803	100
localization:				
cytoplasmic	1191	57.0	143	17.8
cytoplasmic membrane	437	20.9	342	42.6
cell wall	38	1.8	2	0.2
extracellular	56	2.7	53	6.6
unknown	366	17.5	263	32.8



Localization of detected and non-detected proteins of *S. aureus* HG001.

Localization of proteins is depicted according to the PSort annotation. The upper part of the graph indicates numbers of detected proteins, while the lower part shows the numbers on non-detected proteins.

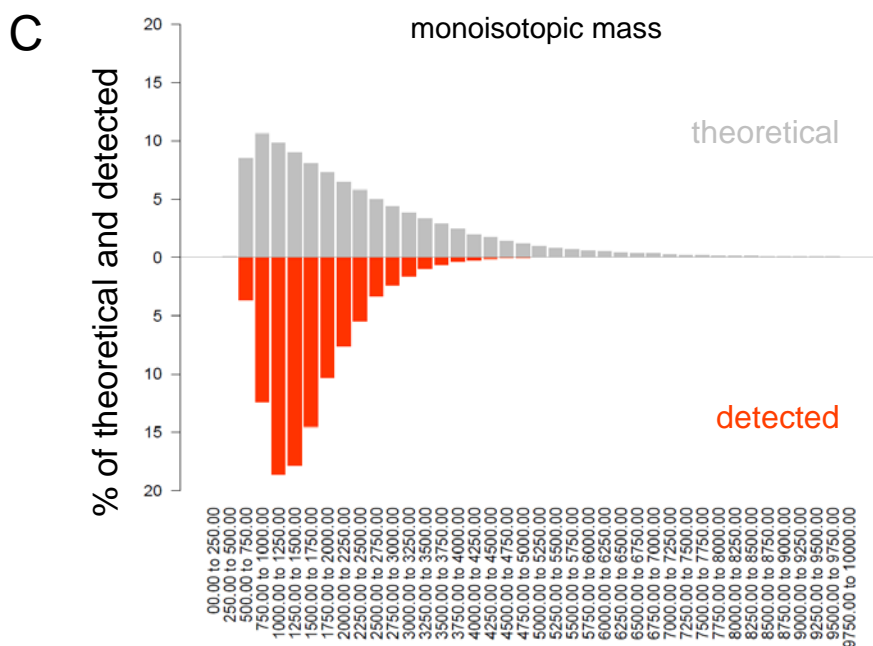
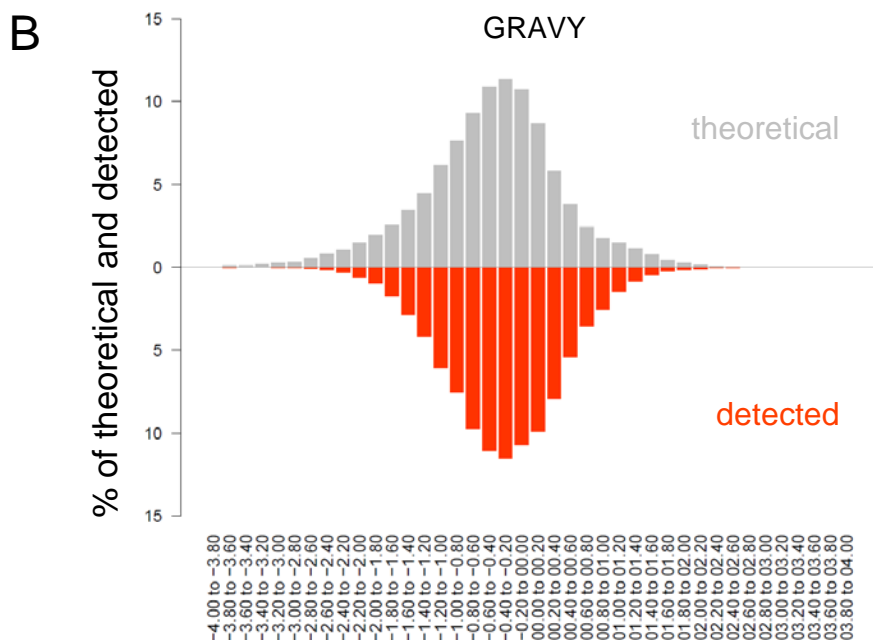
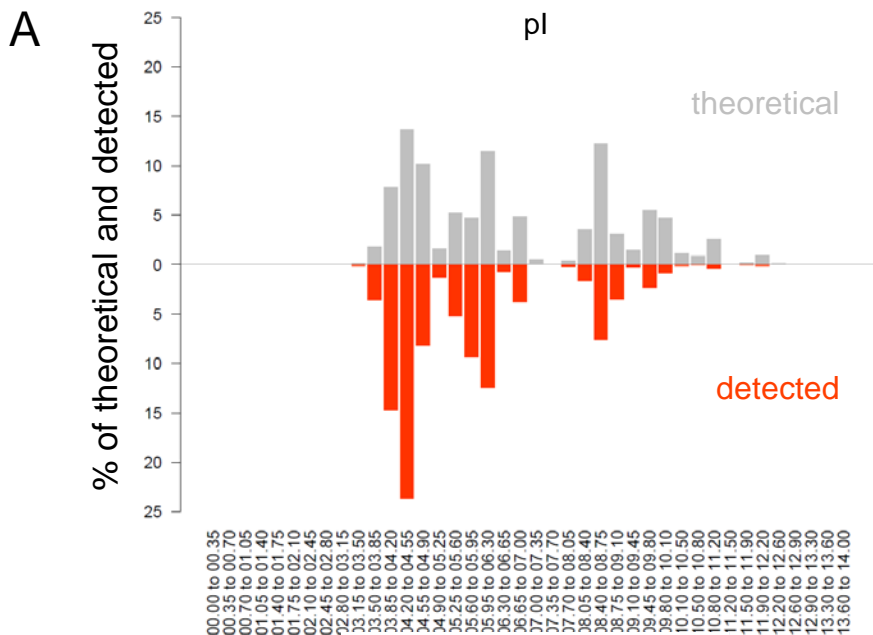
Supplemental Figure S4



Comparison of the distribution of peptide identifications along the pI range for the three methods under investigation.

pI values of detected peptides were exported from the ProteomeCenter tool. The density function and plots were generated using R scripts.

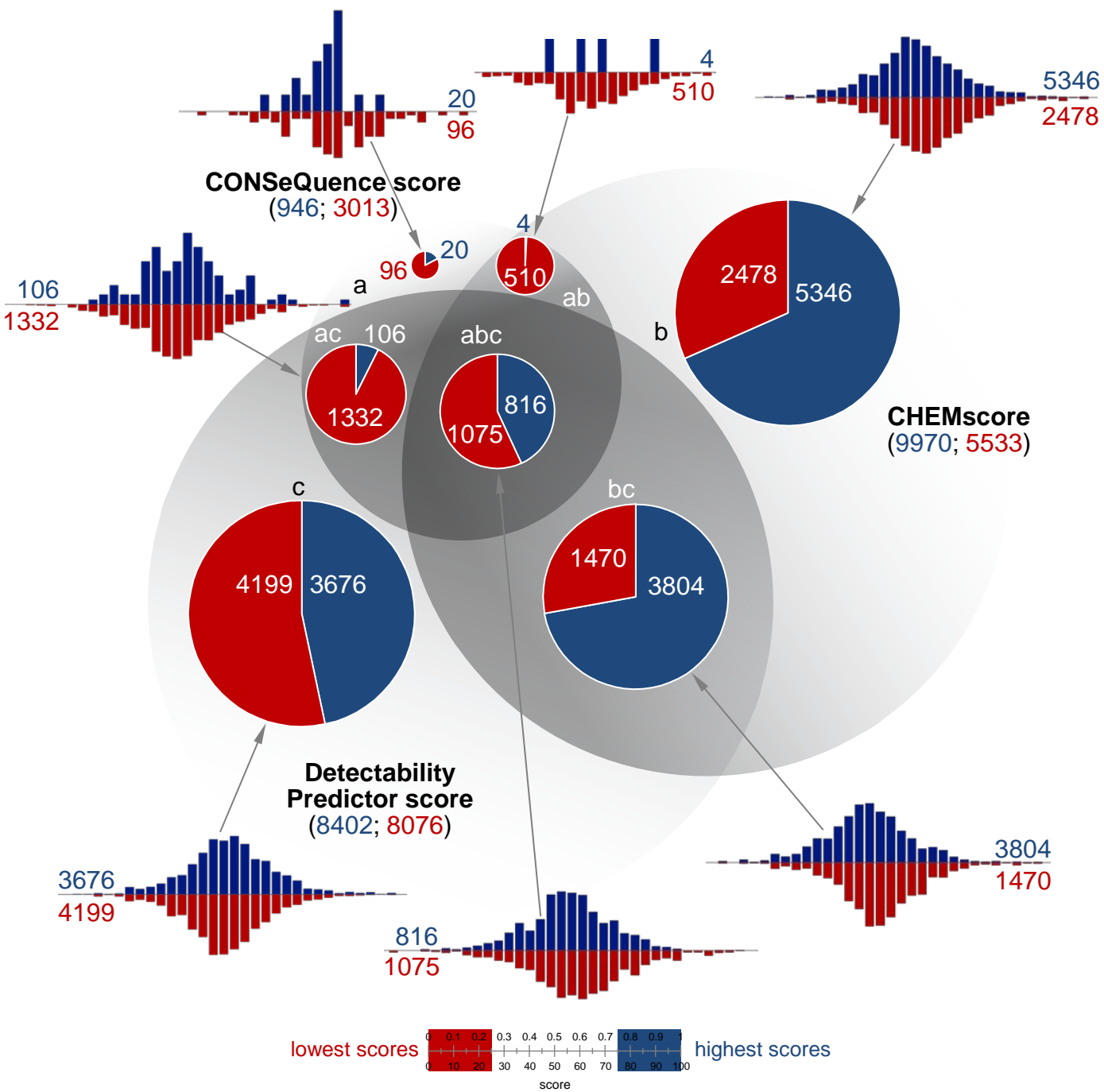
Supplemental Figure S5



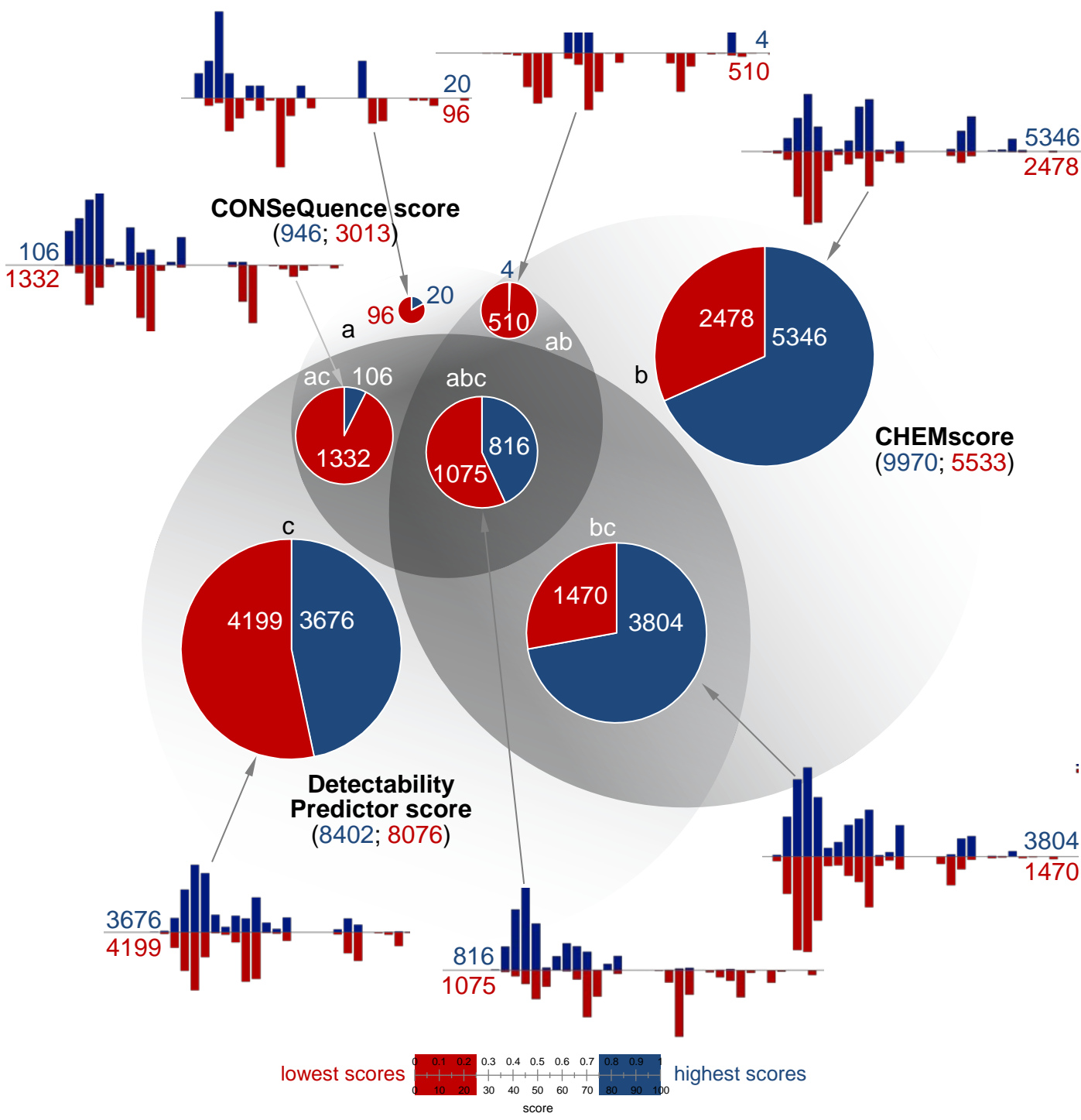
Characterization of the detected peptides and peptides retrieved from an *in silico* digestion by physicochemical properties.

The frequency of peptides is depicted in classes of pI values (A), GRAVY score (B), and monoisotopic mass (C). The upper part of the histograms refers to the set of theoretical peptides from the *in silico* digestion (gray), and the lower part of the histograms displays the data set of detected peptides (orange).

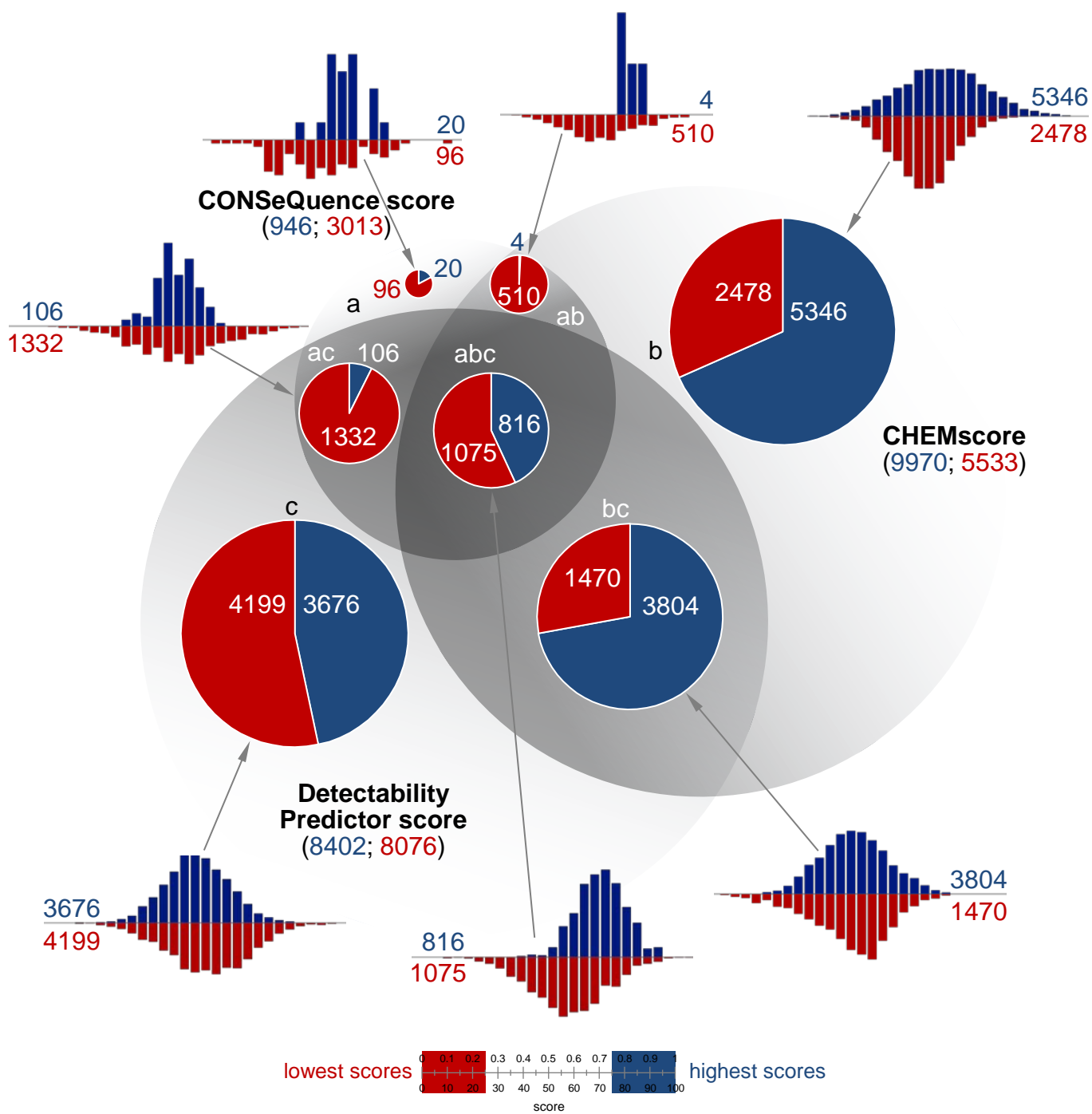
Supplemental Figure S6 A – emPAI –



Supplemental Figure S6 B – p/–



Supplemental Figure S6 C – GRAVY –



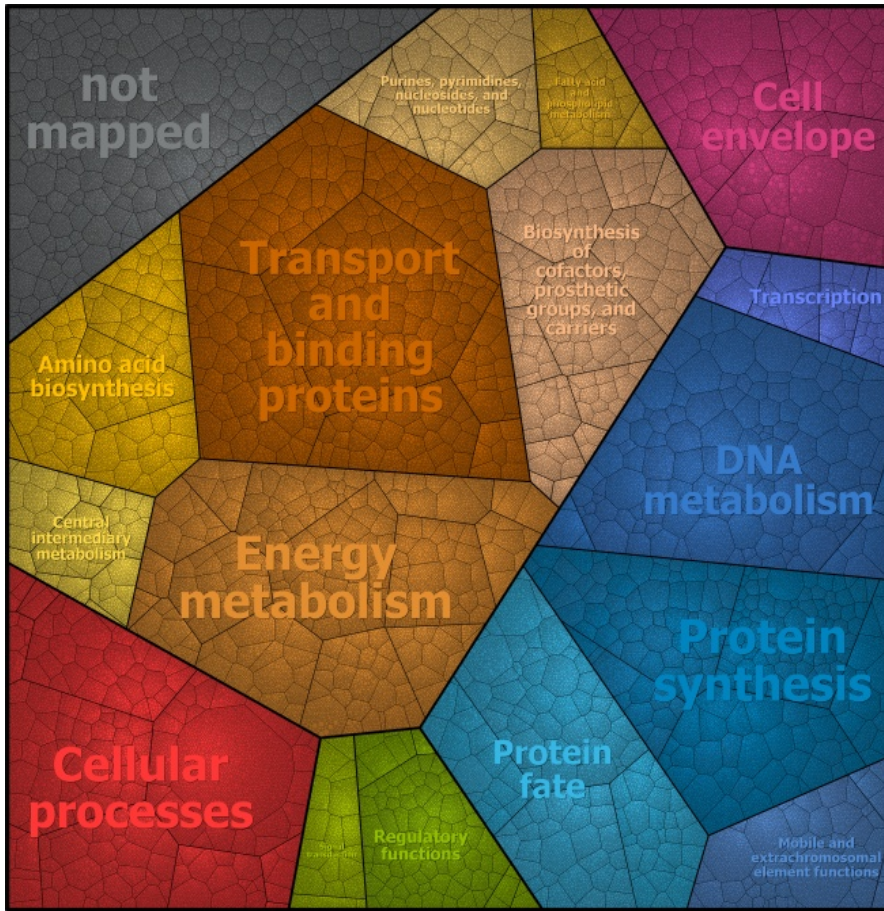
(Supplemental Figure S6)

Comparison of detected peptides with the lowest and the highest score values.

Venn diagrams depict the comparison of peptides with CONSeQuence score > 0.75 , CHEMscore > 75 , and DetectabilityPredictor score > 0.75 , which constitute groups of high score values (colored blue), and of peptides with CONSeQuence score < 0.25 , CHEMscore < 25 , and DetectabilityPredictor score < 0.25 , which were selected as groups of low score values (colored red). Superimposed on the Venn diagram subsets, pie charts correlated by area show the number of peptides with highest and lowest score values. The pie charts are furthermore labeled with the small type characters a (CONSeQuence score), b (CHEMscore), and c (DetectabilityPredictor score) or combinations of them in order to indicate the corresponding Venn diagram intersections. Associated histograms show the distribution and frequency of emPAI (A), pI (B), and GRAVY score (C) values for the subsets of the Venn diagrams. Again, blue indicates the group with highest score values, red marks the group with lowest score values. The x-axis shows categories of values. The y-axis indicates the percentage of values in the corresponding category.

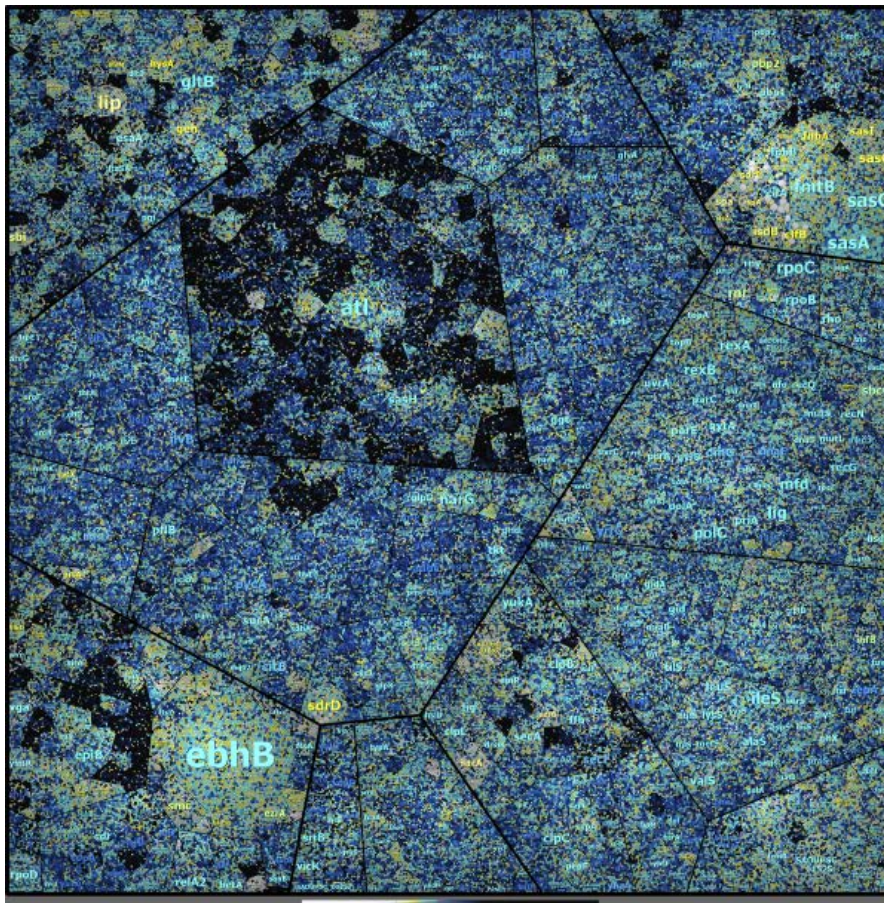
Supplemental Figure S7

A



overview

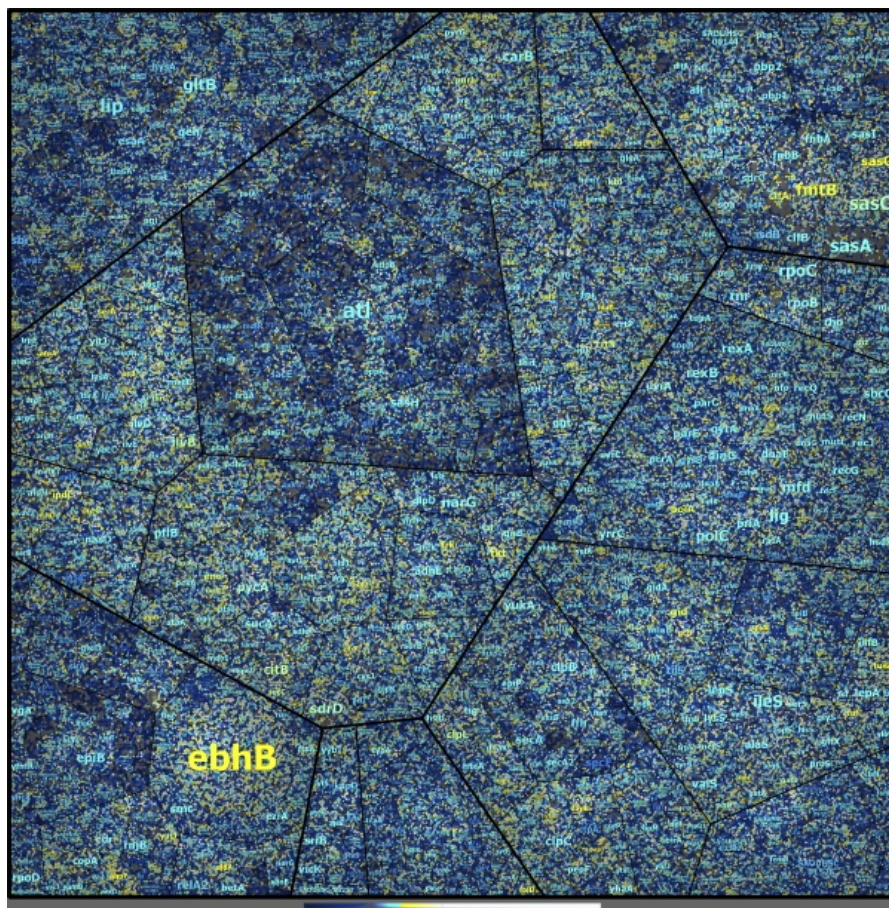
B



GRAVY

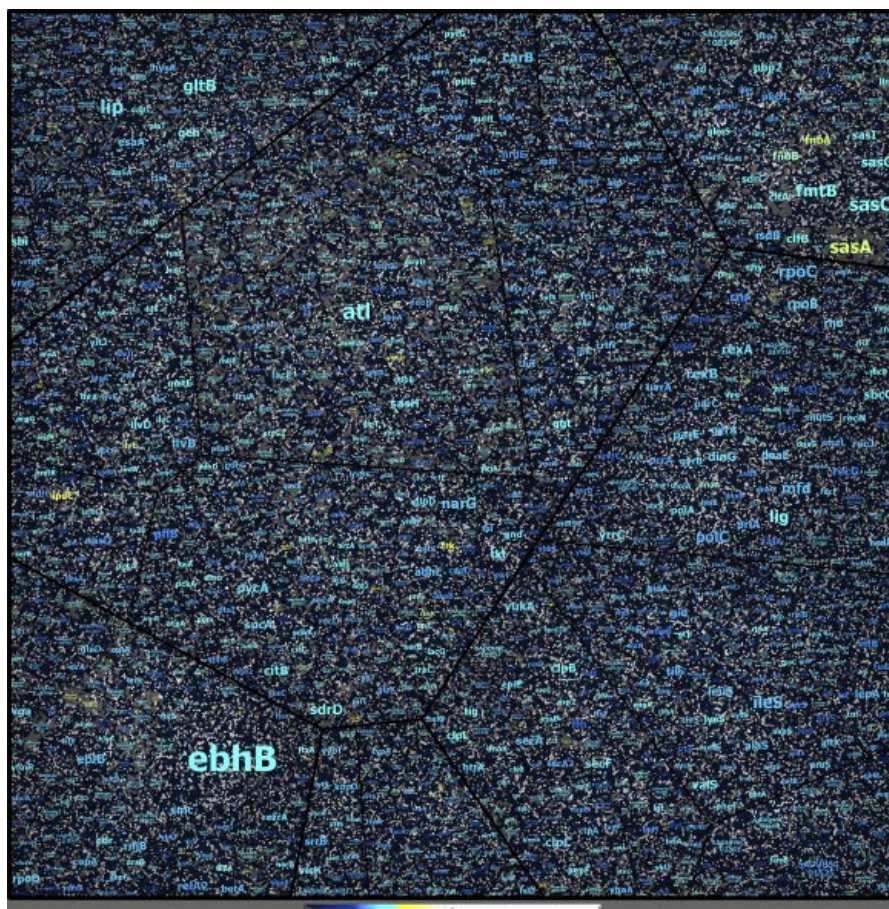
Supplemental Figure S7

C



CONSeQuence
score

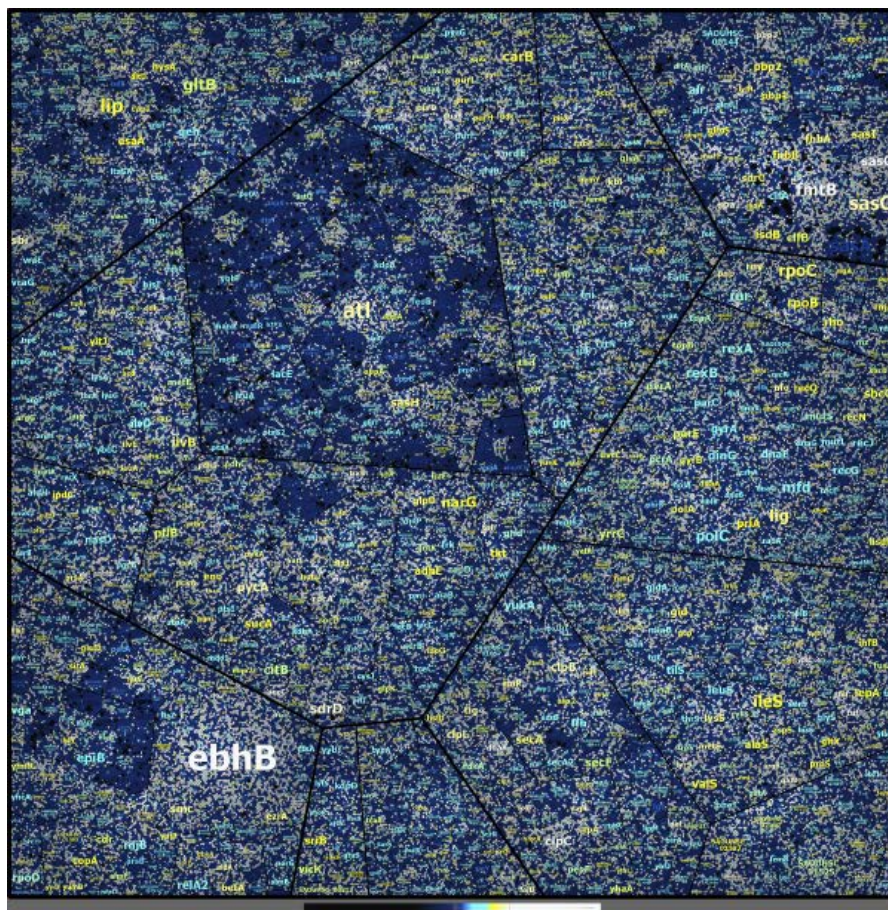
D



CHEMscore

Supplemental Figure S7

E



Detectability
Predictor
score

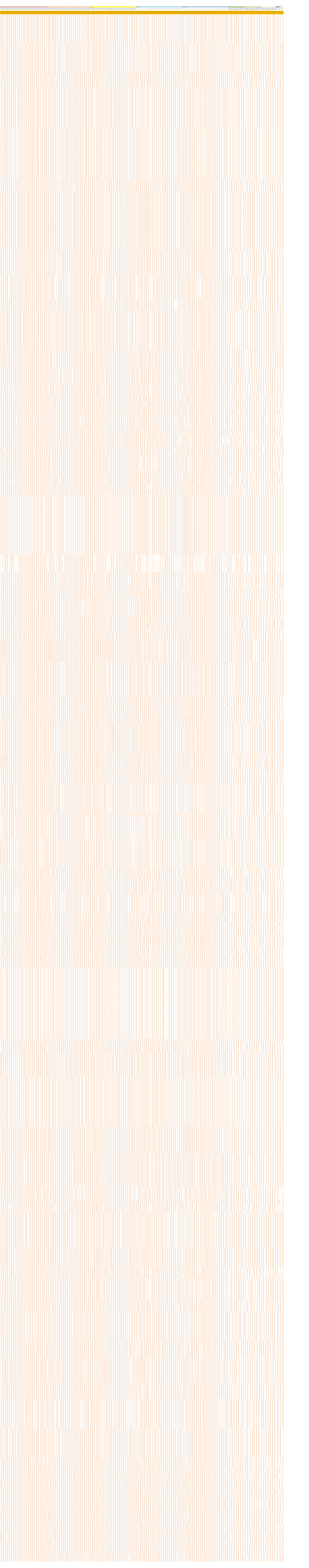
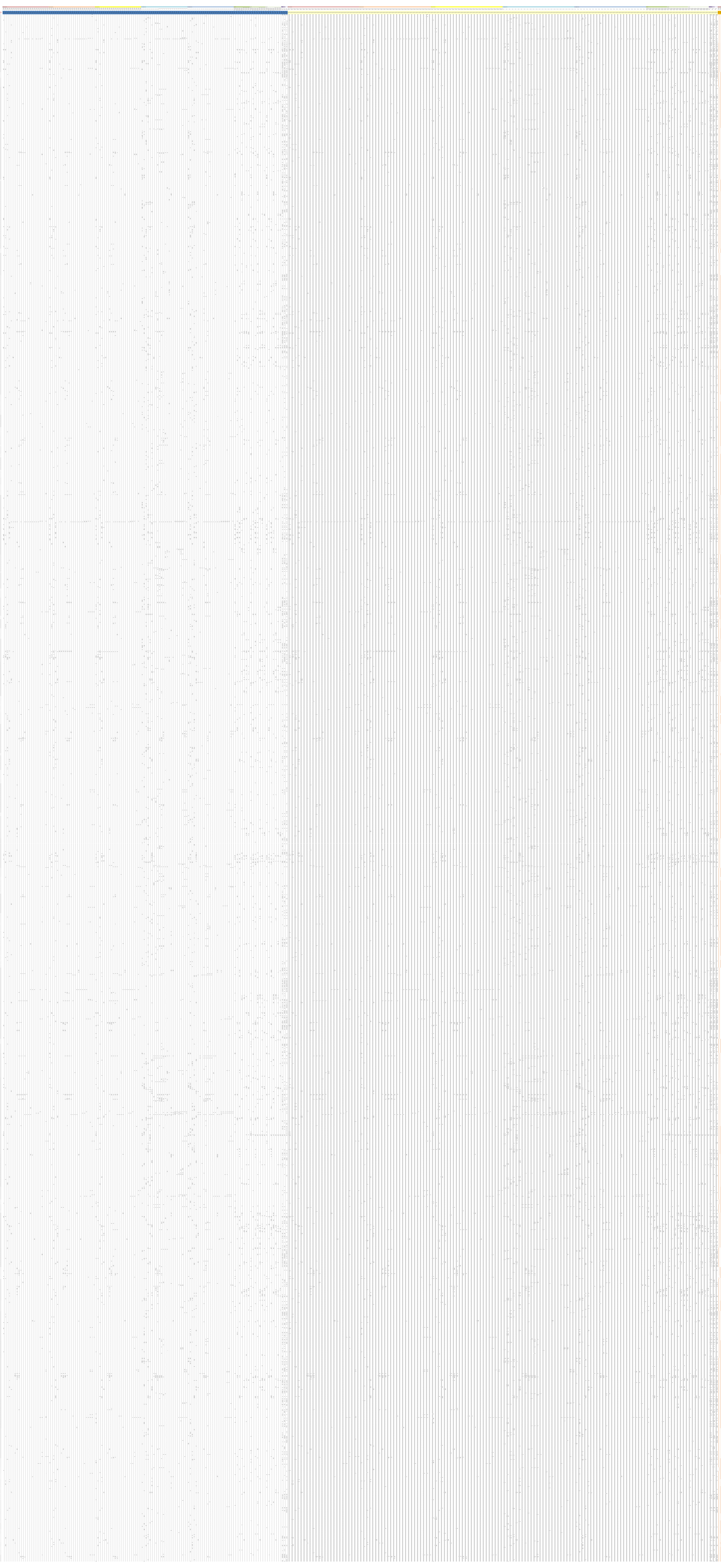
Representation of physicochemical parameters and score values of peptides from an *in silico* digest.

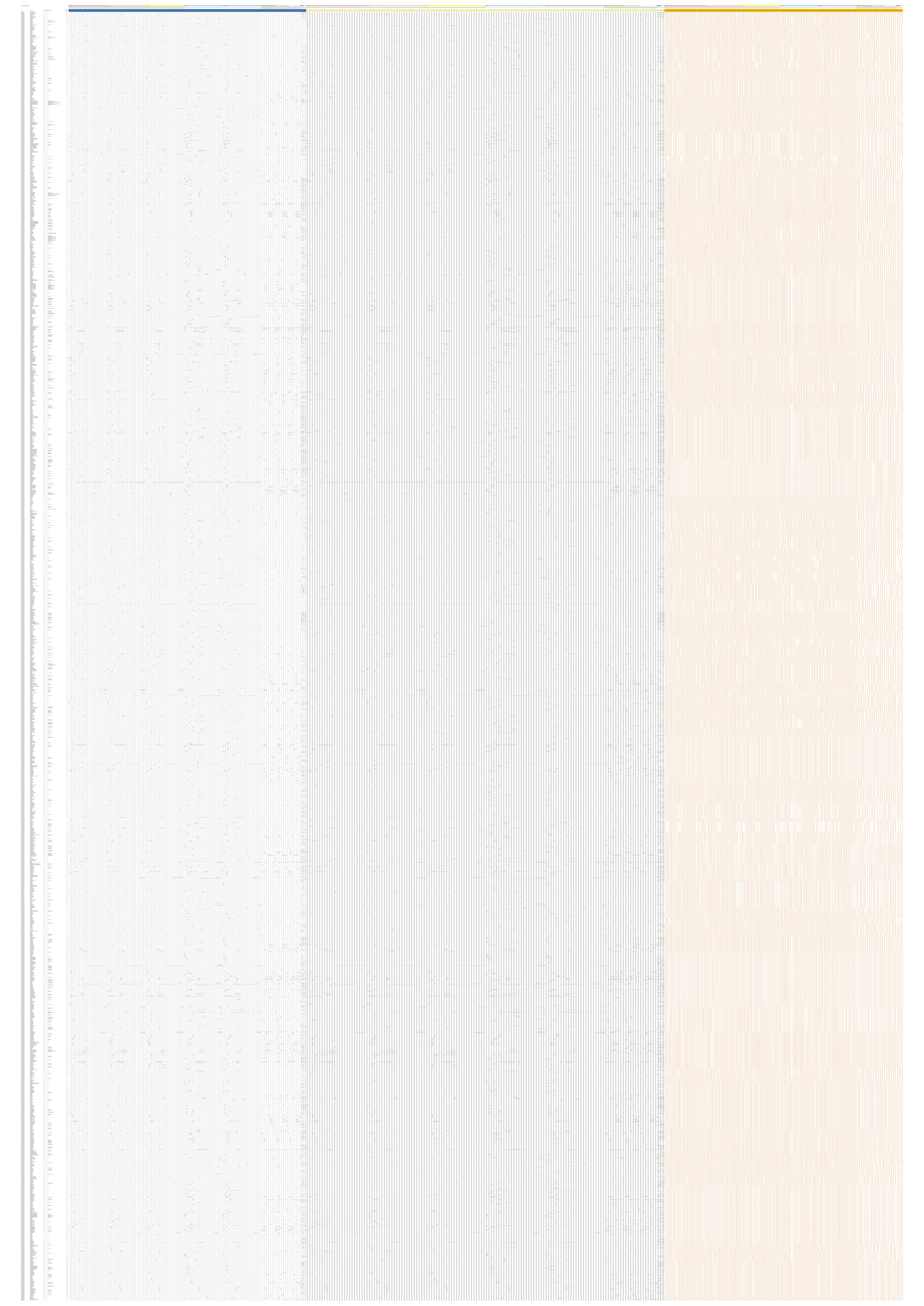
The Voronoi treemap was created on the basis the TIGRFAM protein family classification scheme [Haft et al. (2001) *Nucleic acids research* 29, 41-43] by using HMMER/HMMScan [Finn et al. (2011) *Nucleic acids research* 39, W29-37]. Peptides form the lowest level of area subdivision. The area per peptide represents the peptide length (number of amino acids). Therefore, the area per protein correlates with the protein size. Additionally, the protein label size correlates with the protein size. The overview of the included functional annotations (A) is valid for all parameter-colored Voronoi treemaps. The Voronoi treemap contains about 220,245 theoretically expected peptides from an *in silico* digestion.

Coloring indicates different GRAVY score values (B), CONSeQuence score values (C), CHEMscore values (D), and DetectabilityPredictor score values (E).

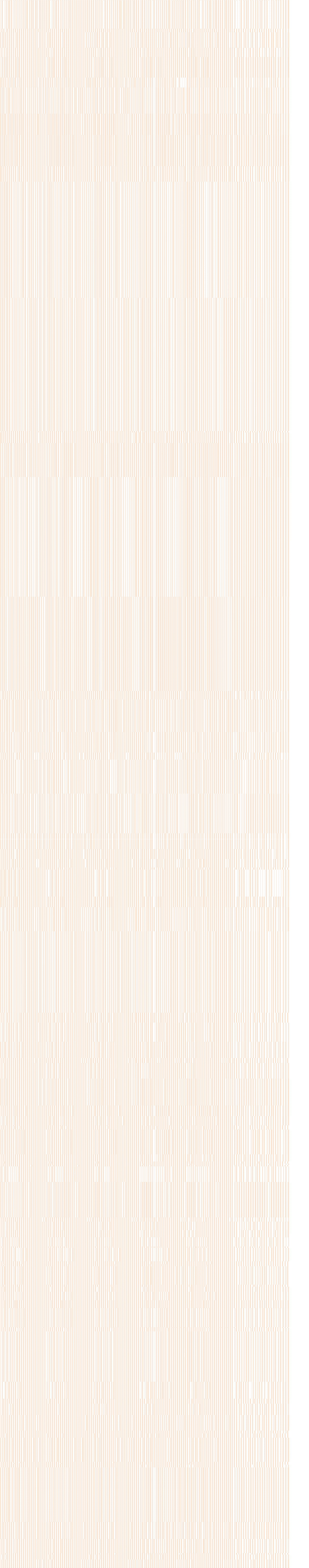
Small GRAVY score values, which indicate less hydrophobic peptides, are colored white, high GRAVY score values, which indicate very hydrophobic peptides, are indicated in dark blue, intermediate GRAVY score values are colored yellow or light blue (B). Missing values are indicated in gray (B).

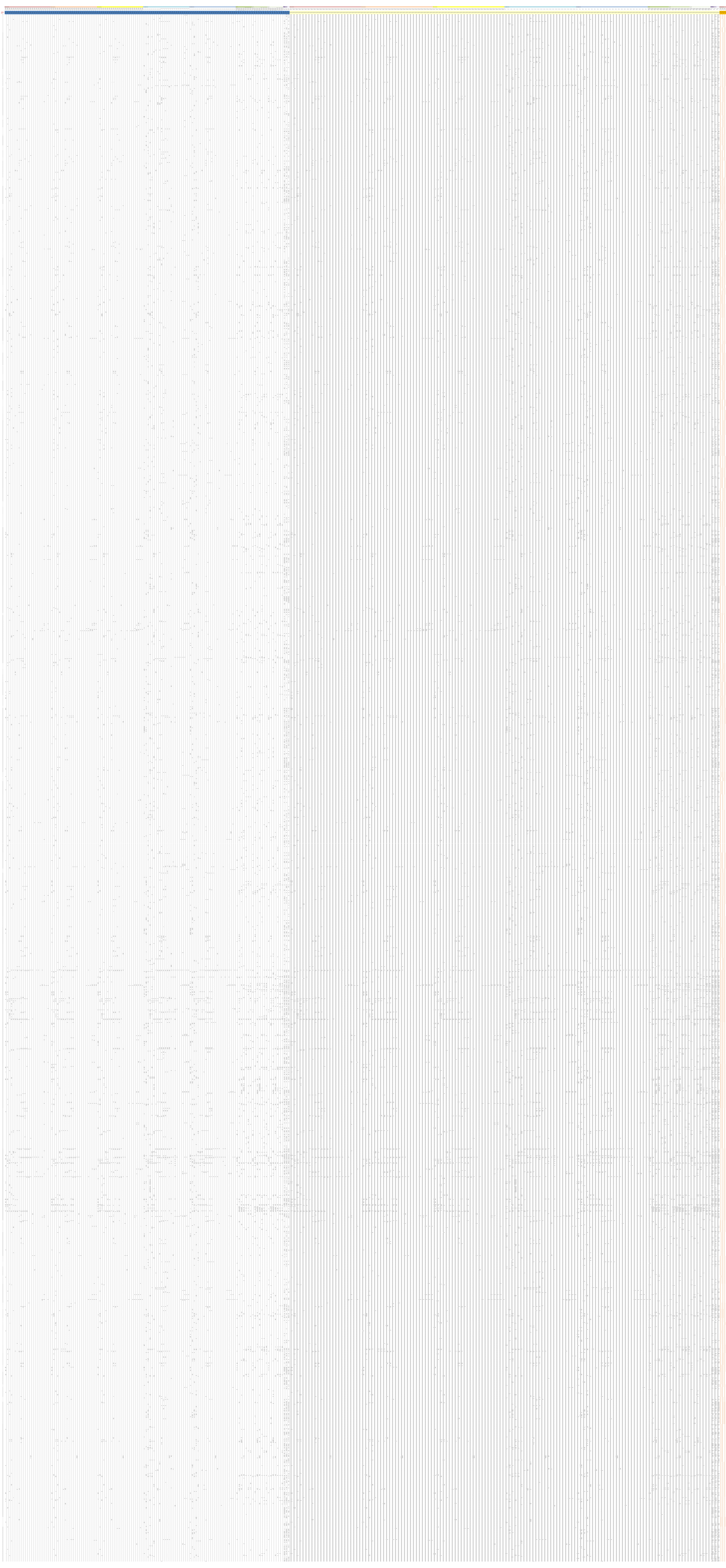
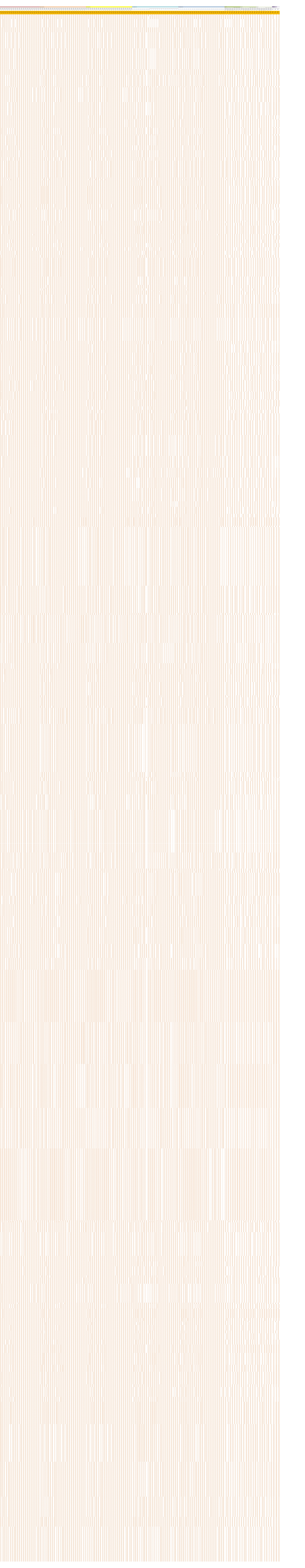
High CONSeQuence score values (C), CHEMscore values (D), and DetectabilityPredictor score values (E) are indicated in white, low score values in dark blue. Intermediate values are colored light blue and yellow (C, D, E). Missing values are indicated in black (C, D, E).

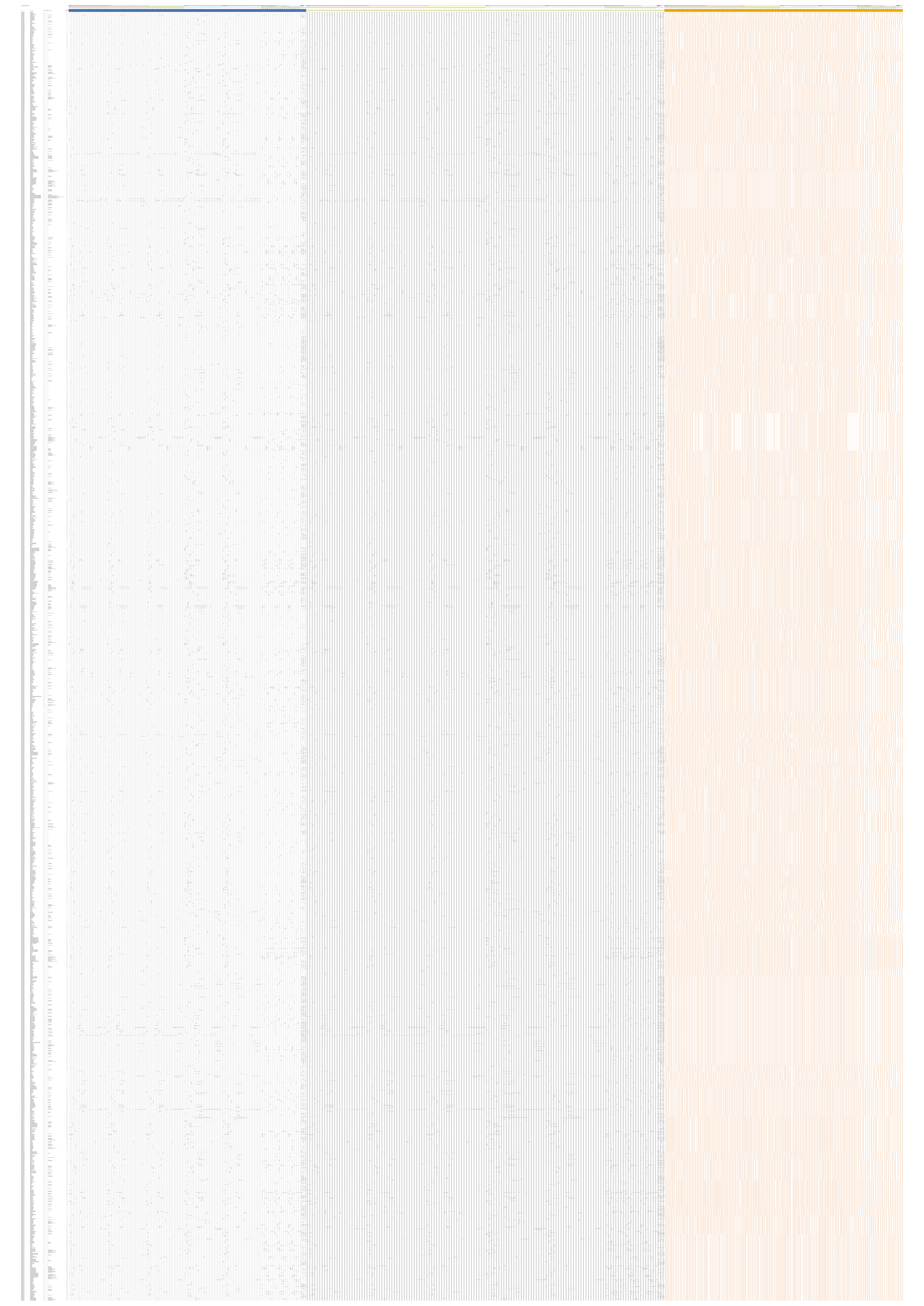




Handwritten text on a grid background, appearing to be a list or index of entries. The text is dense and spans most of the page.

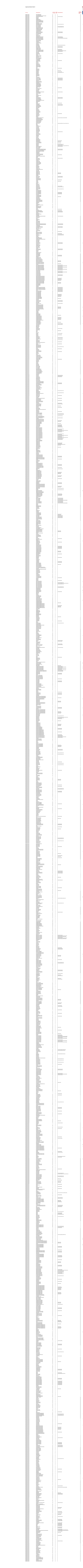
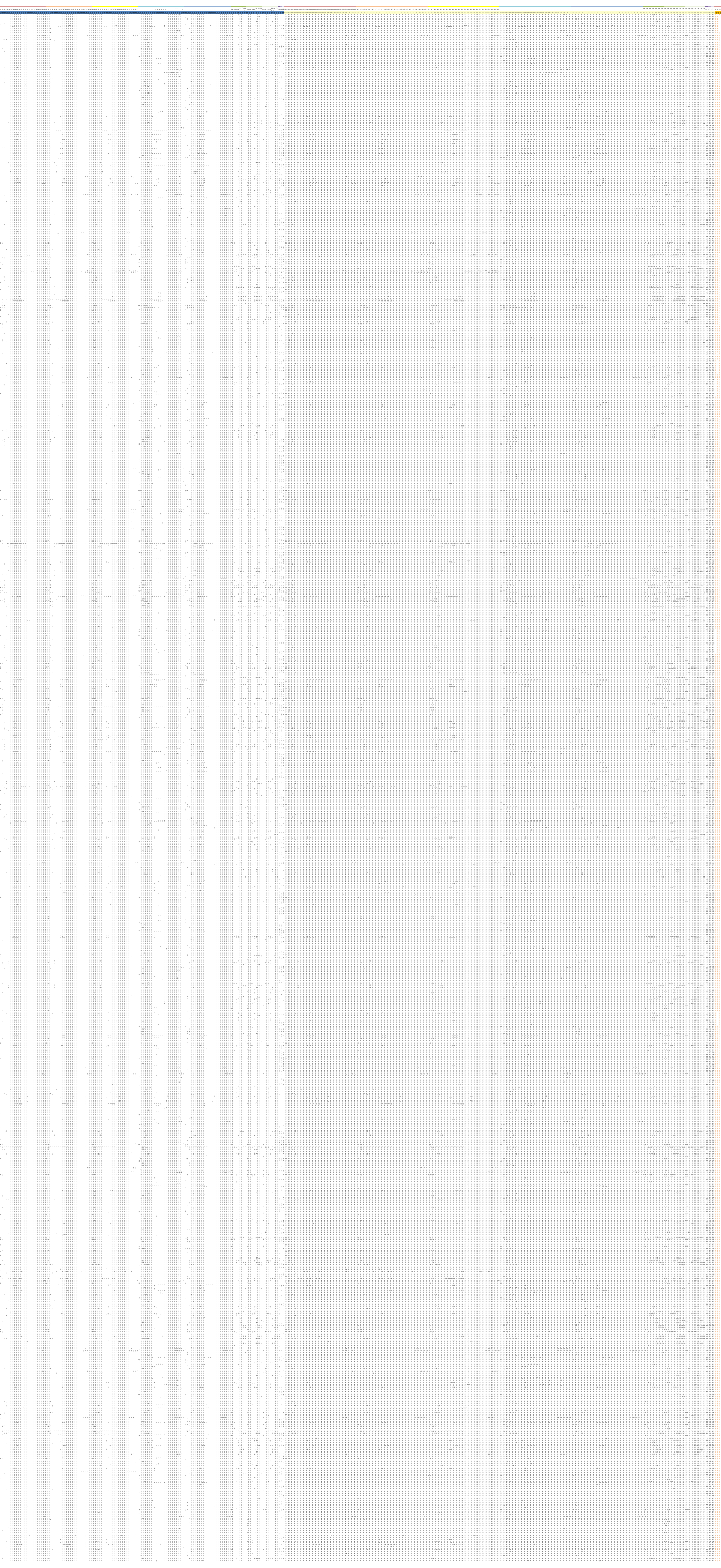


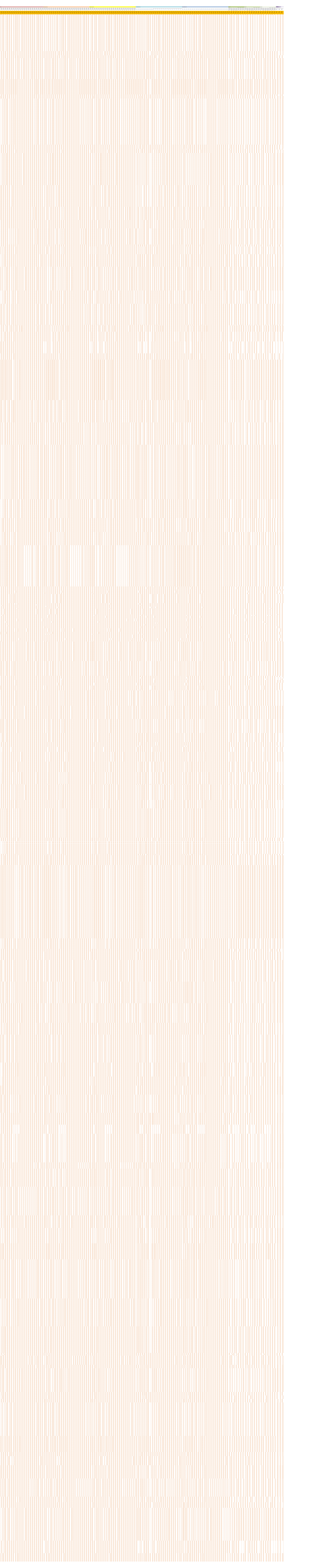
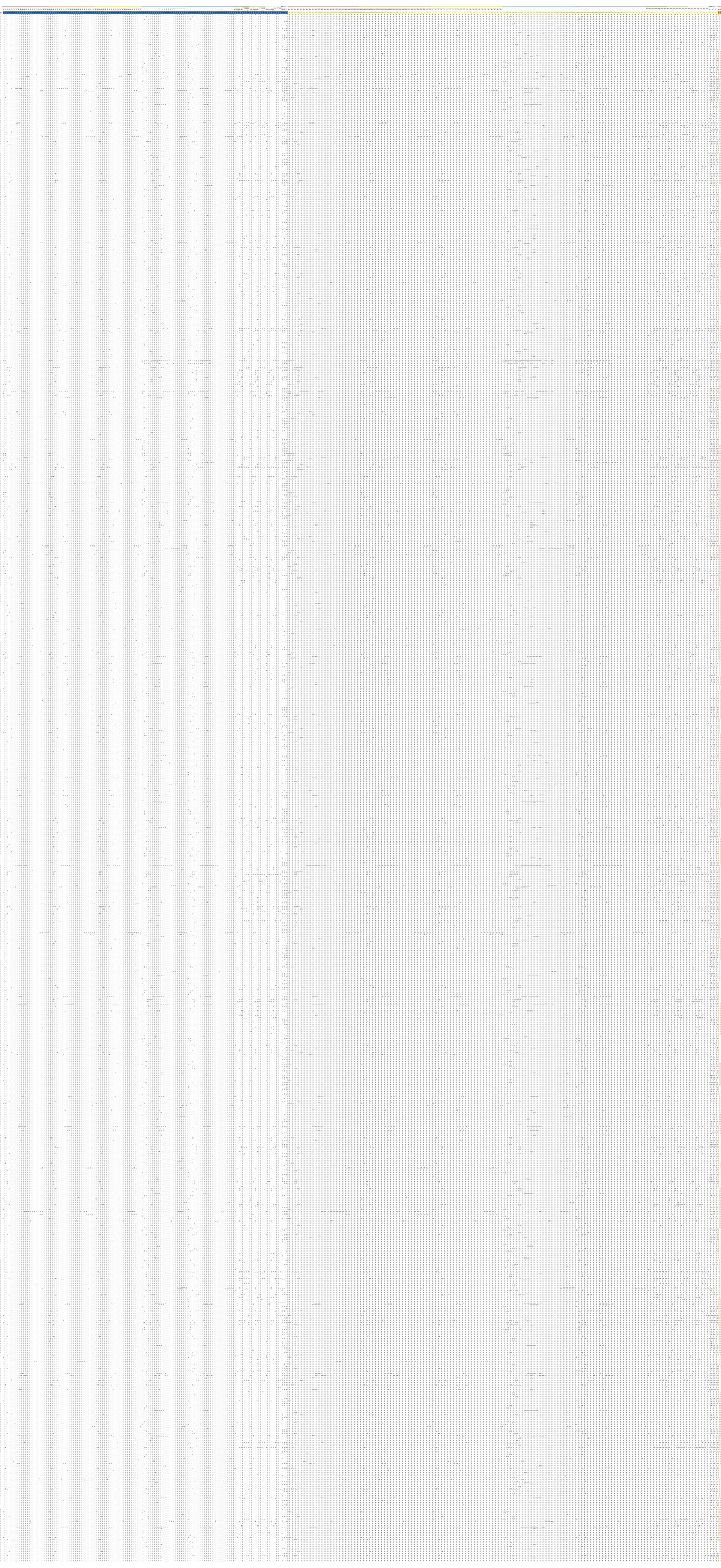


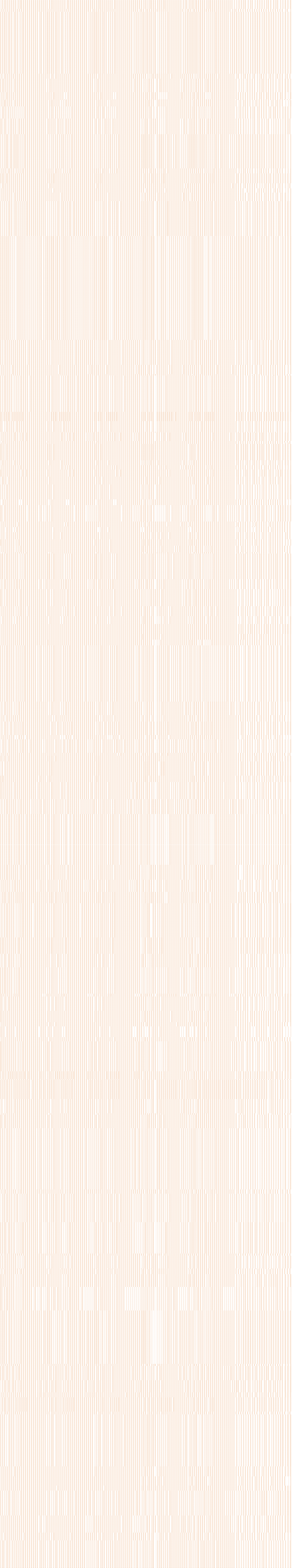
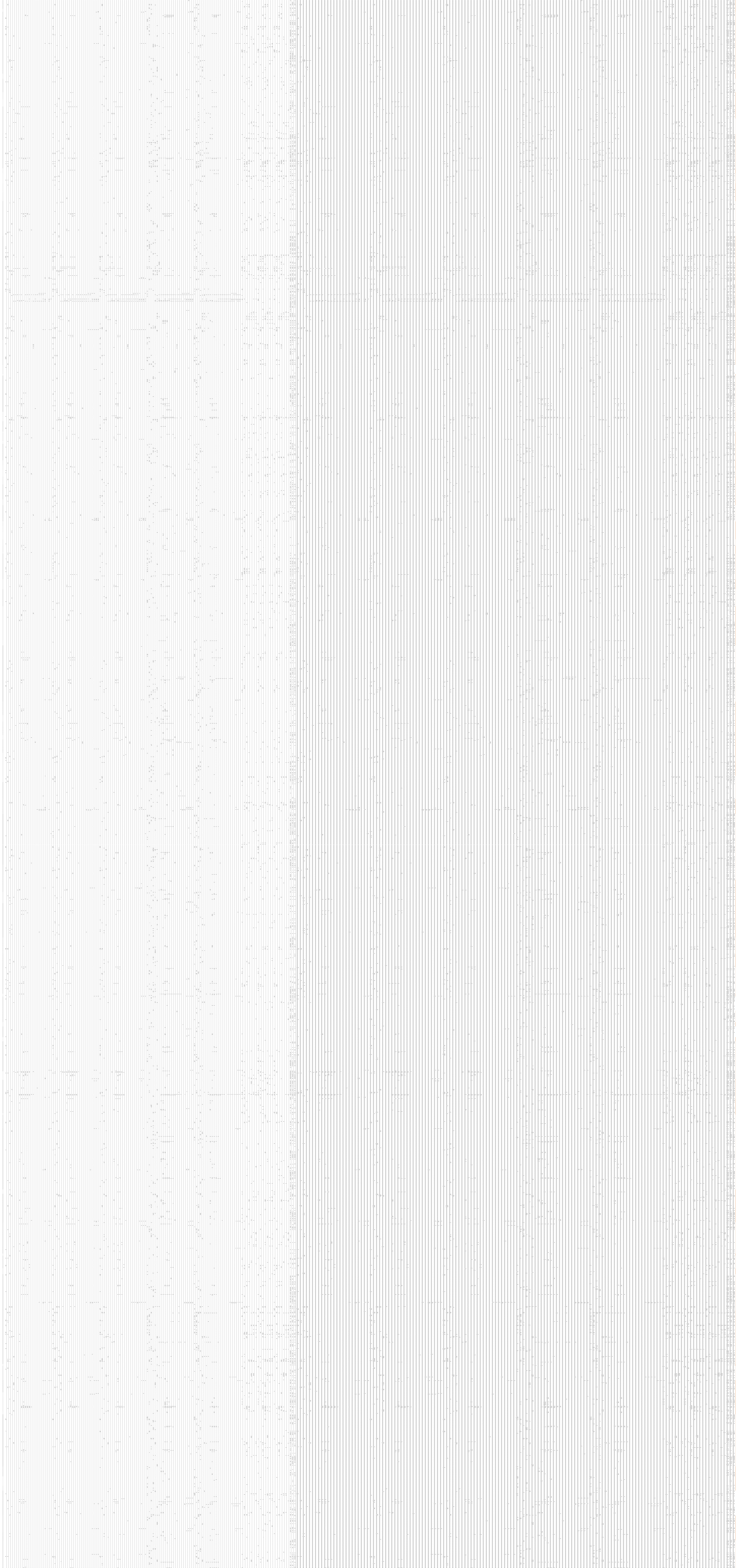


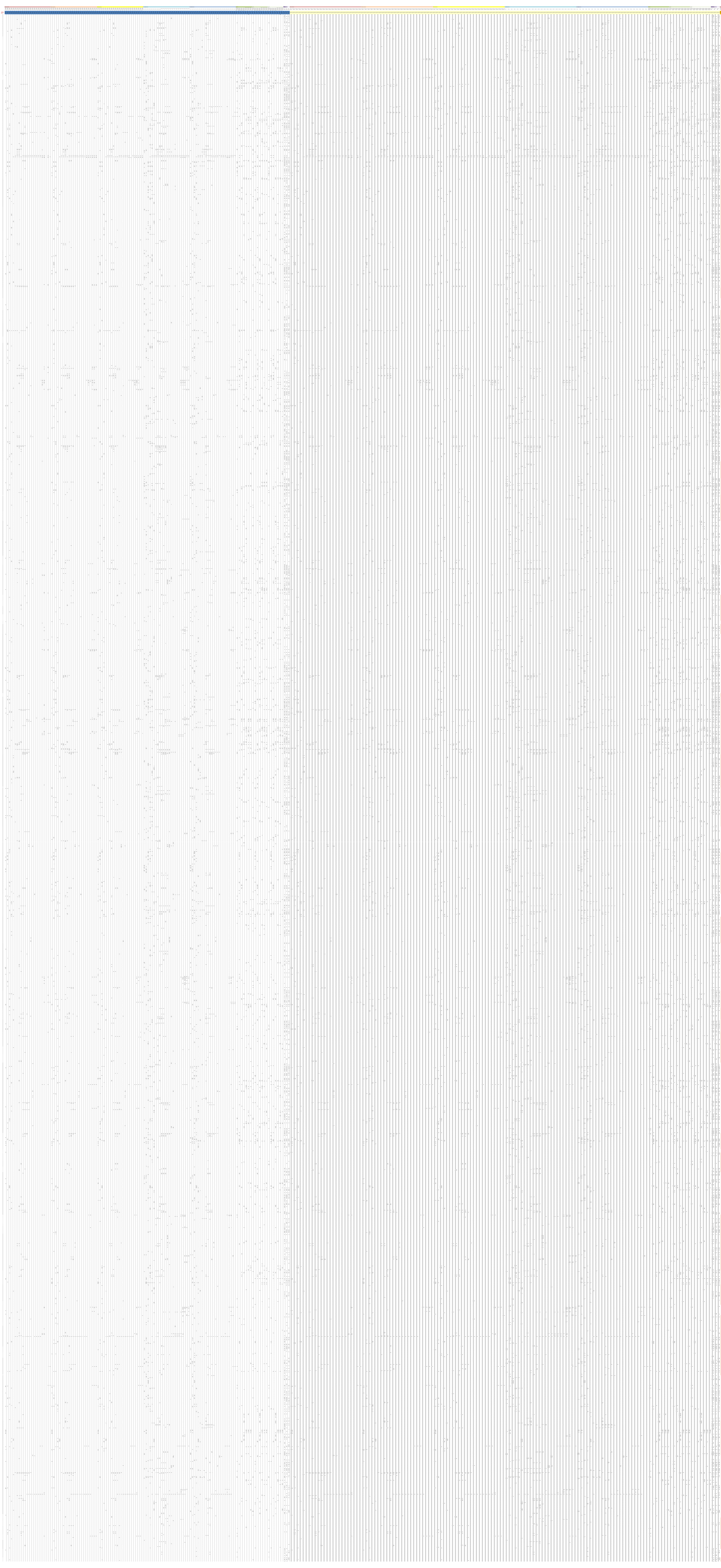
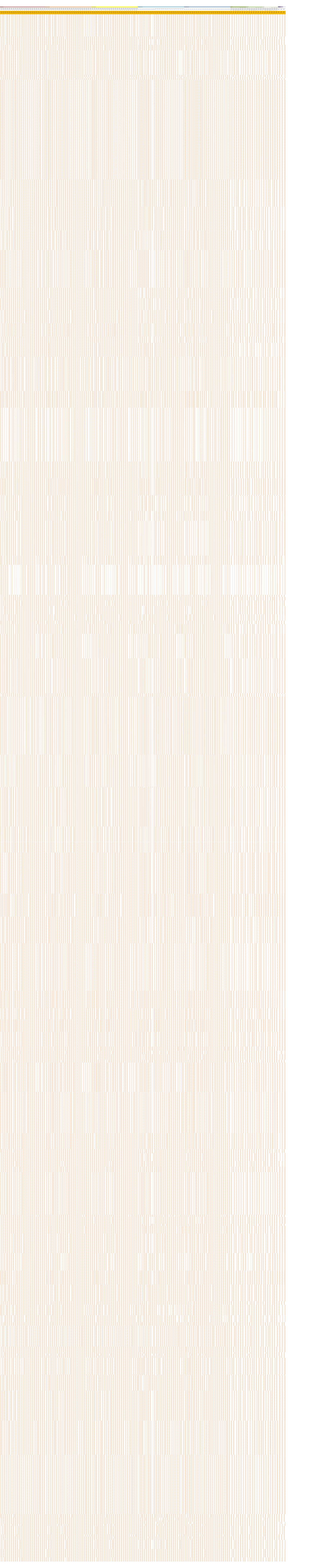




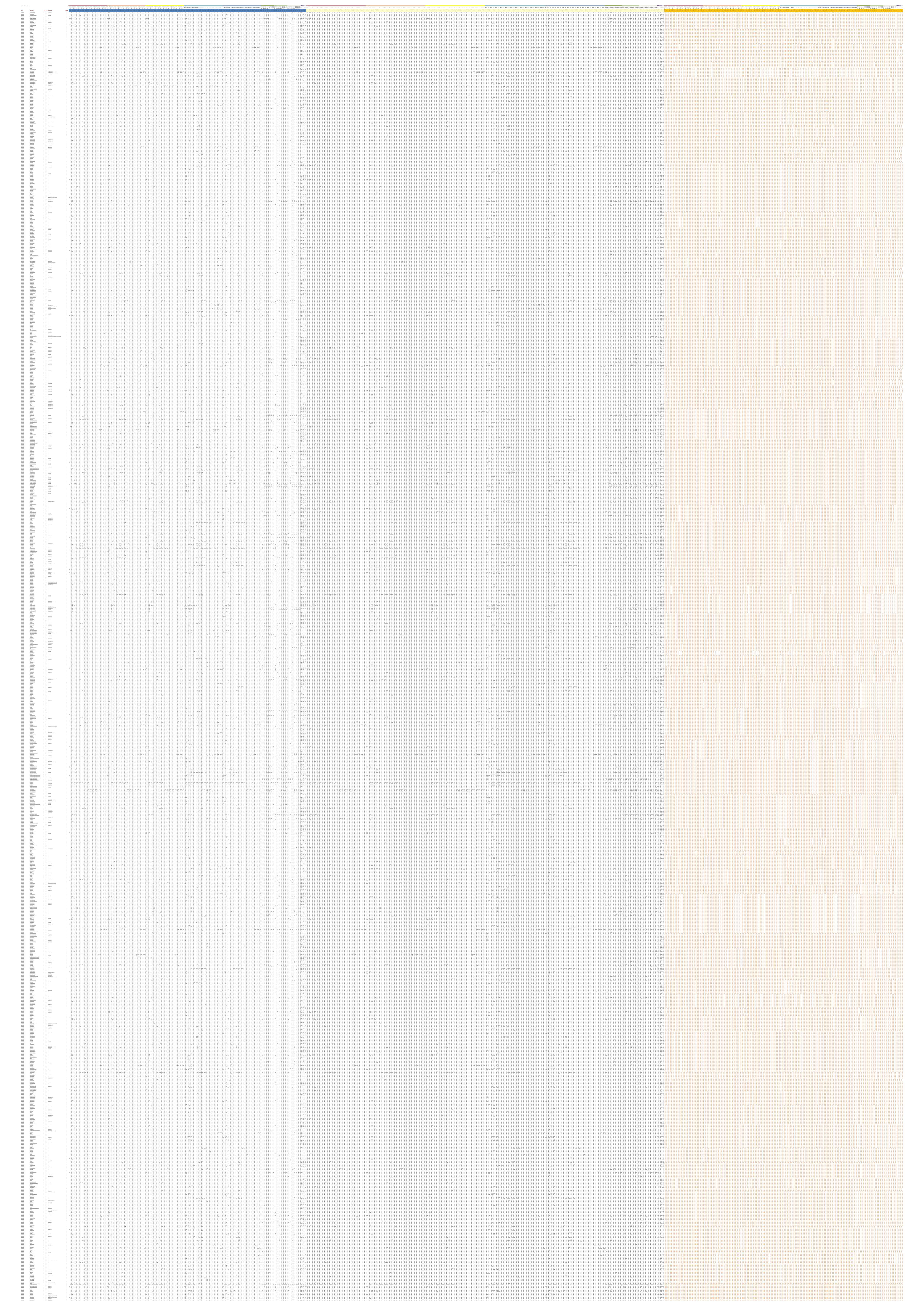




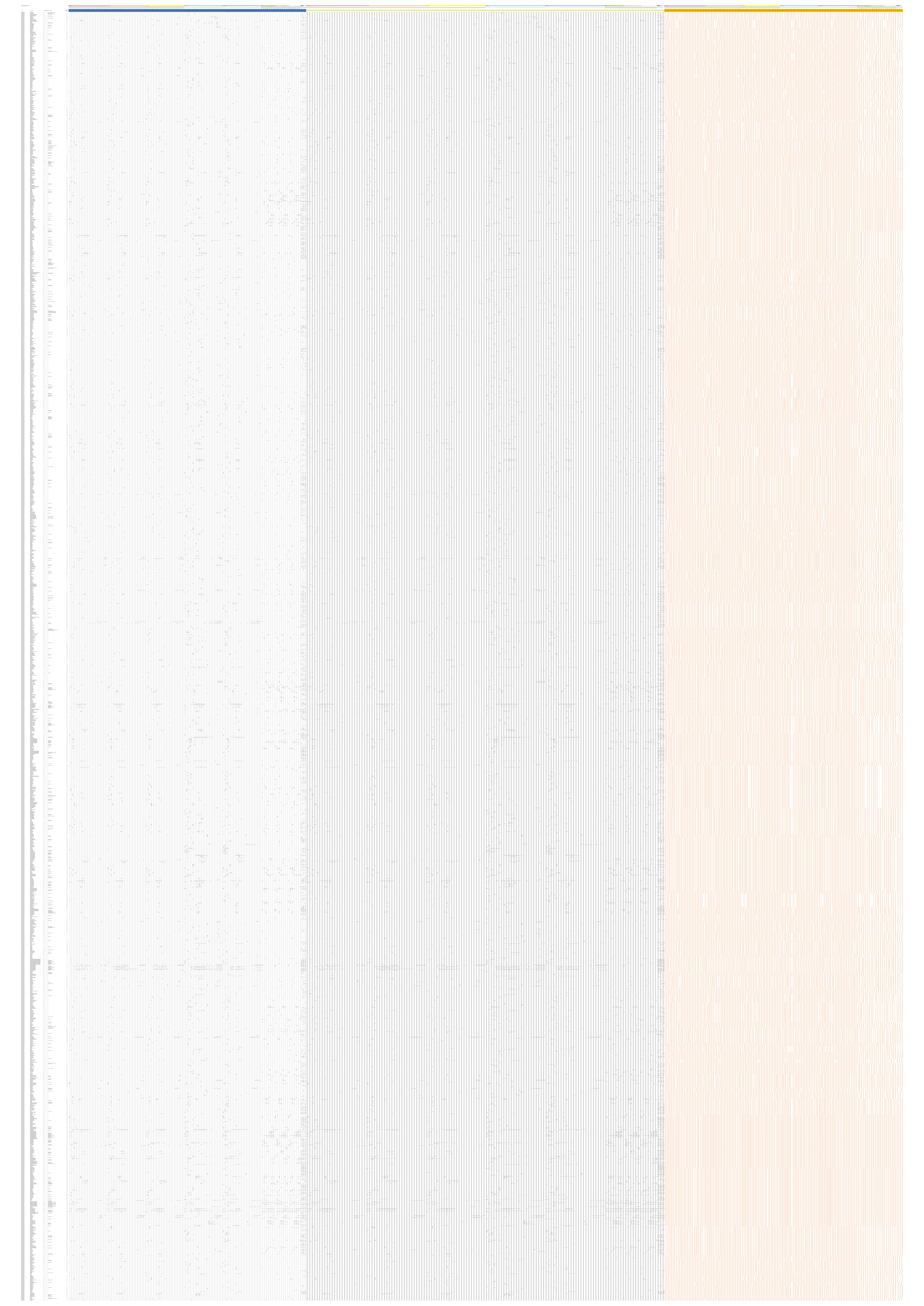












1. The first part of the document is a list of names and their corresponding addresses. The names are listed in the first column, and the addresses are listed in the second column. The names are: John Doe, Jane Smith, and Bob Johnson. The addresses are: 123 Main St, 456 Elm St, and 789 Oak St.

Name	Address
John Doe	123 Main St
Jane Smith	456 Elm St
Bob Johnson	789 Oak St

Name	Address
John Doe	123 Main St
Jane Smith	456 Elm St
Bob Johnson	789 Oak St