EEG Feature search through all electrodes [May 30]

This text continues the results of May 23. The previous goal was to concentrate on the SchRes paper electrodes only. The analysis and last discussion stated to explore all the electrodes over various feature extraction methods.

The hypothesis of structural invariance

The expected result is the list of informative electrodes: the combination (convex or linear) that delivers higher AUC. Since the combination model is under selection now, the hypothesis is. e: The regions with electrodes that carry the informative signal shall generally remain the same regardless of the classification model. We consider the class of model to select from as adequate. So we run an electrode selection algorithm on several models and various extracted features.

Electrode selection algorithm

Since the number of electrodes (four peaks give 512 features for 128 electrodes) is comparable to the number of objects in the dataset (64,...,907) and the EEG ERP electrode signals are highly correlated, the Elastic Net algorithm, used for the feature selection previously, delivers unstable results (see Table 4, Figure 4 in Appendix). Therefore, there are only two solutions for structure optimization 1) a discrete genetic feature selection algorithm and 2) a quadratic programming feature selection algorithm.

Discrete genetic feature selection

The description of the algorithms is in Appendix. The parameters. Number of resulting electrodes: 16 (out of 128), number of the best sets: 3, population size: 14, probability of element mutation: 2/16, iterations: 100 (30 for speed up).

Classifier: logistic regression

Table 1. Top 15 features, electrodes, and peaks, selected after 10.000 populations.

Electrode, peak	Occurencies
A7 LNP	4808
B11 P1	4417
D5 P1	4099
D13 P1	4022
A17 LNP	4022
C30 LNP	3568

D4 LNP	3328
C31 P3	3328
A12 P1	3253
D12 P2	3209
D31 LNP	3139
B32 P1	2943
D5 LNP	2884
D21 P2	2884
D12 P3	2884
D10 P2	2884
C28 P2	2884
A19 P3	2884
D8 P2	2745
A30 LNP	2570
C22 P3	2515

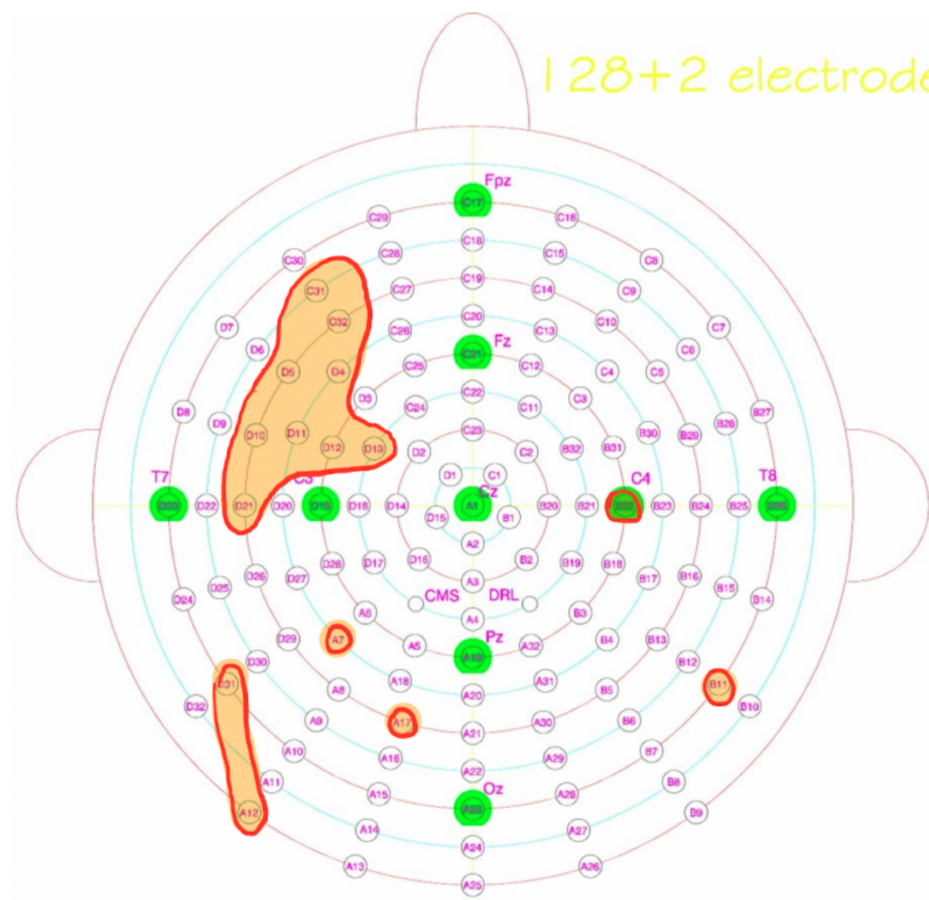

Figure 1.These 15 top most frequent electrodes show areas of signal importance.

Classifier: metric model

Warning! The metric model requires computational resources. Only 30 populations passed. The results below are just an example.

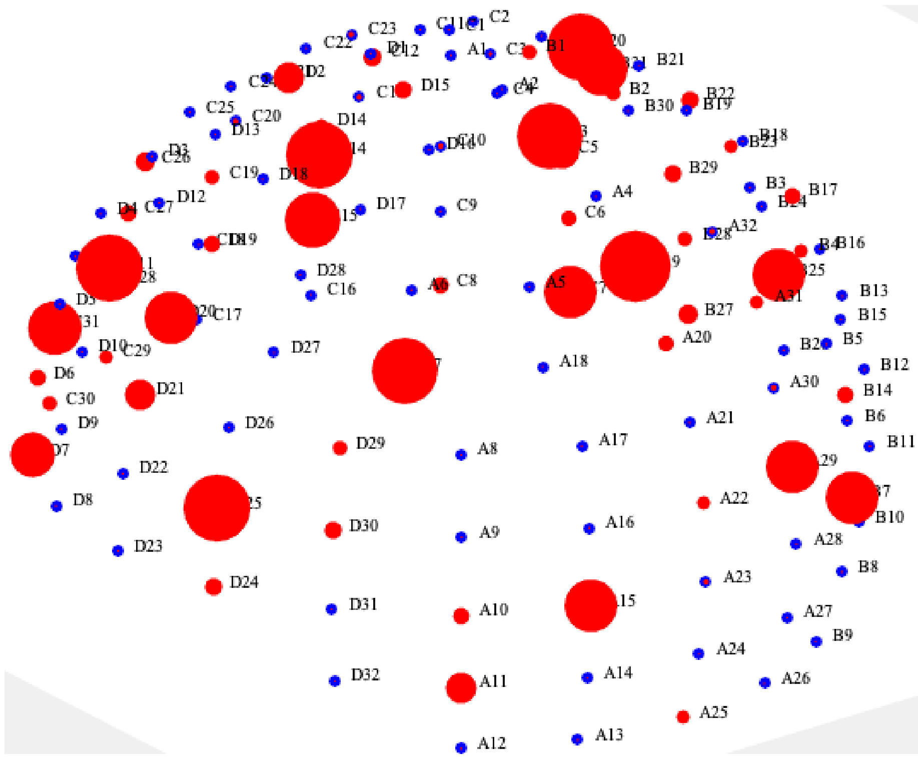

Figure 2. Selected electrodes, three-dimensional placement

Table 2. Most frequent electrodes

Electrode	Score
A19	0.0437
B20	0.0413
C14	0.0413
D11	0.0413
D25	0.0413
A3	0.0406
A7	0.0406
C15	0.0344
C31	0.0331
A15	0.0325
A29	0.0325
B7	0.0325

B25	0.0325
B31	0.0325
C7	0.0325
D20	0.0325

Draft for the later discussion

Also, after the previous discussion, the cross-validation procedure must be carefully selected and declared. Any model selection procedure will lead to overtrained results due to the data complexity and will reduce the sample size. This procedure supposes the train, test, and validation splitting. The new dataset will be fitted by a selected fixed model with parameter fine-tuning. So the cross-validation procedure will be

1) preprocess each user's data individually or align with the other users,
2) keep a fixed proportion of trials for each user in the non-trained set (user balance),
3) shuffle the users and their trials randomly, make the training procedure,
4) (possibly) split the 3) for train and testing to select a model or features,
5) for the 4) case, we can make an average for leave-K-out splitting on model selection,
6) make for the 1) validation average for leave-K-users-out splitting on parameter fine-tuning.

Data cross-validation settings. The proportion for train-test-validation: …; proportion (number of trials per user) for leave-K-user-out: … ; the number of iterations for averaging: …

Warning! Before putting any numbers and proportions here, please analyze the sample size in previous texts. See, for example, Table 7 of Apr 26.

See supplementary technical chart. See Figure 3 for proportion 0.4 of K-leave out cross-validation for 10 folds. It shows how heterogenous the users could be.

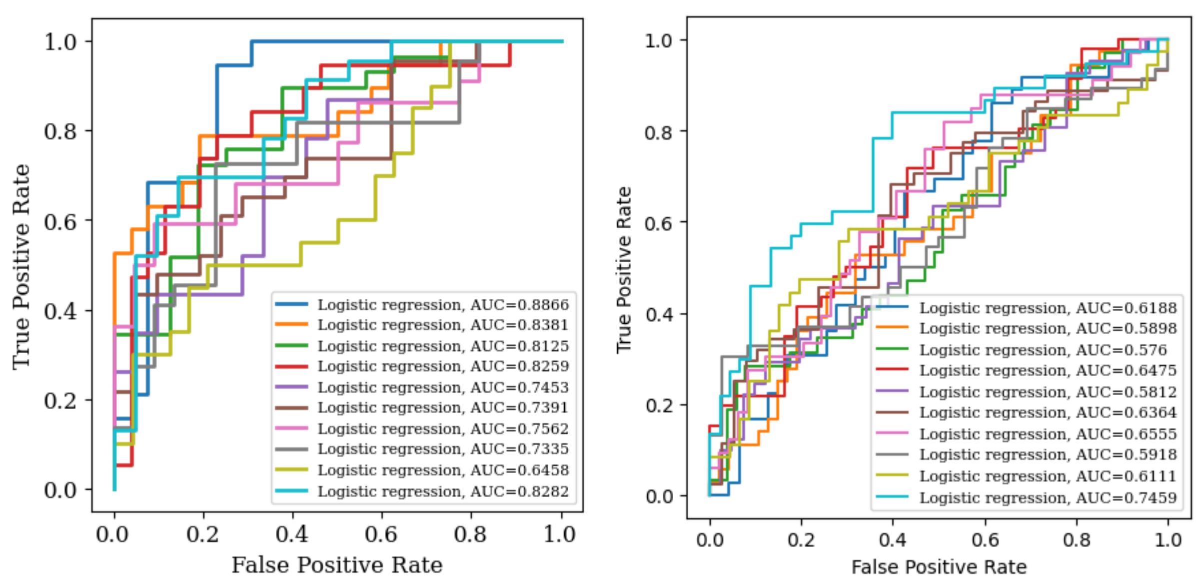

Figure 3. Test ROC for the leave-K-out procedure, user-unbalanced for six selected (left) and all (right) users, and correct responses. K = 10. Test proportion = 0.4.

Discrete genetic algorithm for electrode selection

1. There are set of binary vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_P\}$, $\mathbf{a} \in \{1, \dots, k\}^n$;
2. get two vectors $\mathbf{a}_p, \mathbf{a}_q$, $p, q \in \{1, \dots, P\}$;
3. chose random number $\nu \in \{1, \dots, n-1\}$;
4. split both vectors and change their parts:

$$[a_{p,1}, \dots, a_{p,\nu}, a_{q,\nu+1}, \dots, a_{q,n}] \to \mathbf{a}'_p,$$

$$[a_{q,1}, \dots, a_{q,\nu}, a_{p,\nu+1}, \dots, a_{p,n}] \to \mathbf{a}'_q;$$

5. choose random numbers $\eta_1, \dots, \eta_Q \in \{1, \dots, n\}$;
6. replace values in positions $\eta_1, \dots, \eta_Q$ of the vectors $\mathbf{a}'_p, \mathbf{a}'_q$ for random values from $\{1, \dots, k\}$;
7. repeat items 2-6 $P/2$ times;
8. evaluate the obtained models.

Repeat R times; here P, Q, R are the parameters of the algorithm and k is desired number of categories.

For minor details, check the code.

Feature selection algorithm with L1 regularization

The feature selection algorithms with parameter regularization, used for searching over all electrodes, deliver unstable results. The regularization path significantly varies over slight changes in data due to high feature correlation.

Table 4. Correlation coefficients between parameter vectors along regularization path (rows) for the dataset, split into two subsets

Zero sample subsets intersection	30% sample subsets intersection	60% sample subsets intersection
0.03	0.0	0.33
0.07	0.16	0.25
0.03	0.18	0.26
0.02	0.17	0.21
0.01	0.16	0.2
0.01	0.15	0.19
0.01	0.14	0.18
0.0	0.13	0.19
0.01	0.13	0.18
0.01	0.13	0.22
0.03	0.12	0.23
0.02	0.12	0.26
0.05	0.13	0.26
0.07	0.11	0.29
0.13	0.15	0.29

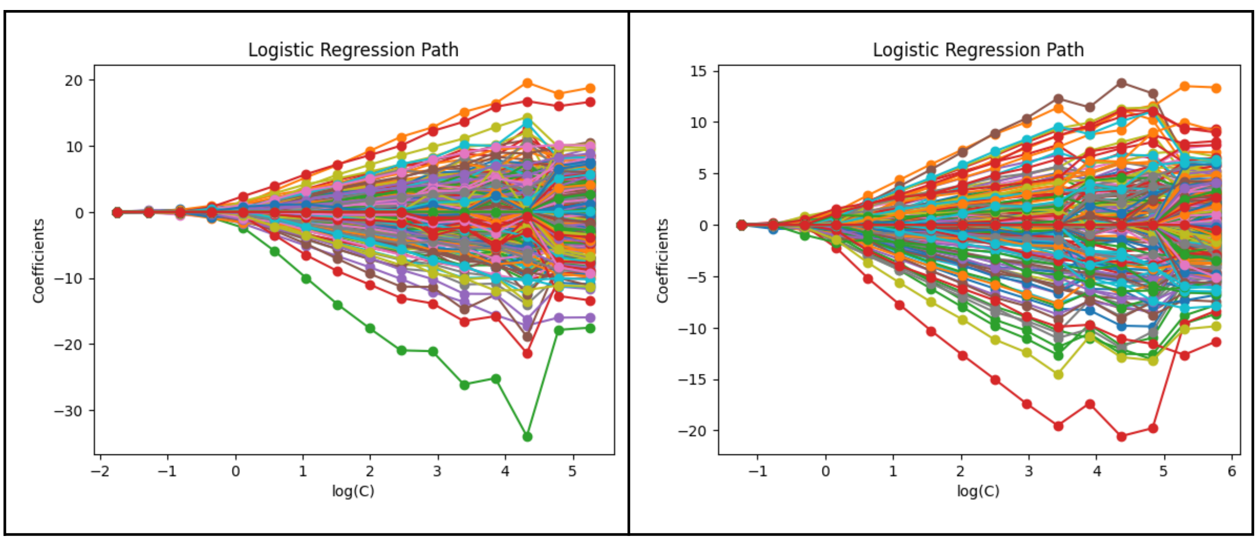

Figure 4. Regularization path for two subsets

References

1. Use the xdf format instead of mat https://mne.tools/dev/auto_examples/io/read_xdf.html
2. Draw the graph by the coordinates import networkx
3. Install Ubuntu on Mac mini
4. The open-source project https://github.com/intheon/freud
5. Grid search for model selection one, two, feature selection ridge, lasso

Genetic feature selection over logistic regression

Each peak was averaged, the time series were standardized, and six selected users.

Table 3. Full list from Table 1.

('A7 LNP', 4808.0)
('B11 P1', 4417.0)
('D5 P1', 4099.0)
('D13 P1', 4022.0)
('A17 LNP', 4022.0)
('C30 LNP', 3568.0)
('D4 LNP', 3328.0)
('C31 P3', 3328.0)
('A12 P1', 3253.0)
('D12 P2', 3209.0)
('D31 LNP', 3139.0)
('B32 P1', 2943.0)
('D5 LNP', 2884.0)
('D21 P2', 2884.0)
('D12 P3', 2884.0)
('D10 P2', 2884.0)
('C28 P2', 2884.0)
('A19 P3', 2884.0)
('D8 P2', 2745.0)
('A30 LNP', 2570.0)
('C22 P3', 2515.0)
('A20 P1', 1737.0)
('B7 P1', 1662.0)
('A2 P3', 1630.0)
('D15 LNP', 1624.0)
('D9 P3', 1607.0)
('C29 P2', 1607.0)
('D10 P1', 1603.0)
('A31 LNP', 1584.0)
('D27 P3', 1563.0)
('C14 P3', 1480.0)
('D29 P2', 1355.0)
('B5 P2', 1321.0)
('B8 P1', 1300.0)
('D32 P2', 1294.0)
('B27 P3', 1294.0)
('D9 P1', 1293.0)
('C4 P3', 1293.0)
('C32 LNP', 1293.0)
('C3 P2', 1293.0)
('B5 P1', 1293.0)
('B10 P1', 1293.0)
('A31 P3', 1293.0)
('A16 P3', 1293.0)
('D18 P3', 1278.0)
('C10 LNP', 1277.0)
('B9 P1', 1277.0)
('B25 LNP', 1277.0)
('B18 P2', 1277.0)
('B11 P3', 1277.0)
('A7 P3', 1277.0)
('A23 P2', 1277.0)
('D8 P1', 1238.0)
('D11 P1', 1238.0)
('C4 P1', 1238.0)
('C11 P1', 1238.0)
('A13 P2', 1238.0)
('A10 LNP', 1238.0)
('A15 LNP', 1216.0)
('C11 P3', 1215.0)
('A1 LNP', 1139.0)
('D16 P1', 1138.0)
('C5 P1', 1138.0)
('C31 LNP', 1138.0)
('C15 P2', 1138.0)
('B4 LNP', 1138.0)
('B31 P3', 1138.0)
('B25 P2', 1138.0)
('A15 P2', 1138.0)
('A10 P3', 1138.0)
('B22 P2', 813.0)
('C24 P1', 644.0)
('A5 LNP', 644.0)
('B29 LNP', 578.0)
('C5 LNP', 577.0)
('D24 LNP', 521.0)
('C9 P2', 521.0)
('D12 P1', 446.0)
('C16 P1', 446.0)
('D6 P3', 444.0)
('D2 P1', 444.0)
('D19 P2', 444.0)
('B30 P2', 444.0)
('B18 LNP', 444.0)
('A24 P1', 444.0)
('B13 P2', 414.0)
('C14 P2', 392.0)
('A21 P2', 370.0)
('D20 P1', 369.0)
('C3 P3', 369.0)
('C17 P3', 369.0)
('B31 P2', 369.0)
('B24 LNP', 369.0)
('B15 LNP', 369.0)
('A23 LNP', 369.0)
('D23 P3', 363.0)
('A4 P3', 356.0)
('D14 P2', 348.0)
('A1 P2', 348.0)
('A3 P1', 338.0)
('D7 P3', 337.0)
('D3 P1', 337.0)

('C6 P2', 337.0)
('B24 P1', 337.0)
('A5 P3', 337.0)
('A14 P1', 337.0)
('A13 LNP', 337.0)
('C11 P2', 327.0)
('D32 LNP', 325.0)
('D2 LNP', 325.0)
('C7 LNP', 325.0)
('C32 P3', 325.0)
('C25 P1', 325.0)
('C23 P1', 325.0)
('C12 LNP', 325.0)
('B3 P2', 325.0)
('B28 P1', 325.0)
('B16 P1', 325.0)
('C27 P2', 323.0)
('A29 P1', 317.0)
('A21 LNP', 309.0)
('A18 P2', 308.0)
('D22 P1', 307.0)
('D14 P3', 307.0)
('C25 P2', 307.0)
('C19 P1', 307.0)
('B25 P3', 307.0)
('B24 P3', 307.0)
('A18 P3', 307.0)
('A16 P1', 307.0)
('A30 P2', 241.0)
('D24 P2', 240.0)
('D22 P3', 240.0)
('D15 P2', 240.0)
('C30 P3', 240.0)
('C16 P3', 240.0)
('B3 P1', 240.0)
('A10 P1', 240.0)
('A22 P3', 100.0)
('C19 P2', 95.0)
('D19 LNP', 78.0)
('D2 P3', 77.0)
('C7 P2', 77.0)
('C4 LNP', 77.0)
('B9 P2', 77.0)
('B12 P1', 67.0)
('A21 P1', 28.0)
('D13 LNP', 26.0)
('C30 P1', 26.0)
('C23 P2', 26.0)
('B17 P1', 26.0)
('A9 P1', 26.0)
('A3 P2', 26.0)
('A29 P2', 26.0)
('A26 P1', 26.0)
('A17 P2', 26.0)
('D28 P1', 23.0)
('D21 P1', 23.0)
('C9 P1', 23.0)
('C9 LNP', 23.0)
('C28 P3', 23.0)
('C17 LNP', 23.0)
('B22 P1', 23.0)
('A25 LNP', 23.0)
('C22 P1', 16.0)
('D8 P3', 15.0)
('D6 P1', 15.0)
('D30 P1', 15.0)
('D17 P1', 15.0)
('C23 LNP', 15.0)
('C20 LNP', 15.0)
('C1 LNP', 15.0)
('A27 P1', 15.0)
('A2 P1', 15.0)
('A10 P2', 15.0)
('B20 LNP', 3.0)
('A21 P3', 3.0)
('D29 P3', 2.0)
('D23 P2', 2.0)
('D18 P1', 2.0)
('D1 P3', 2.0)
('D1 P1', 2.0)
('C25 LNP', 2.0)
('C24 LNP', 2.0)
('C17 P1', 2.0)
('C15 LNP', 2.0)
('B26 P1', 2.0)
('B18 P1', 2.0)
('B17 P3', 2.0)
('B11 LNP', 2.0)
('A6 LNP', 2.0)
('A30 P3', 2.0)
('A29 LNP', 2.0)
('A27 LNP', 2.0)
('A23 P3', 2.0)
('A18 P1', 2.0)
('A12 LNP', 2.0)
('A11 LNP', 2.0)
('D9 P2', 1.0)
('D7 LNP', 1.0)
('D30 LNP', 1.0)
('D3 P2', 1.0)
('D25 P3', 1.0)
('D16 P3', 1.0)
('C7 P3', 1.0)
('C6 P1', 1.0)
('C29 P3', 1.0)
('C27 LNP', 1.0)
('C21 P1', 1.0)
('C19 P3', 1.0)
('C18 P1', 1.0)
('C1 P3', 1.0)
('B30 P1', 1.0)
('B27 LNP', 1.0)
('B26 LNP', 1.0)
('B13 LNP', 1.0)
('A30 P1', 1.0)
('A22 P2', 1.0)
('A22 P1', 1.0)
('A19 P1', 1.0)
('A16 P2', 1.0)
('A16 LNP', 1.0)
('D9 LNP', 0.0)
('D8 LNP', 0.0)
('D7 P2', 0.0)
('D7 P1', 0.0)
('D6 P2', 0.0)
('D6 LNP', 0.0)
('D5 P3', 0.0)
('D5 P2', 0.0)
('D4 P3', 0.0)
('D4 P2', 0.0)
('D4 P1', 0.0)
('D32 P3', 0.0)
('D32 P1', 0.0)
('D31 P3', 0.0)
('D31 P2', 0.0)
('D31 P1', 0.0)
('D30 P3', 0.0)
('D30 P2', 0.0)
('D3 P3', 0.0)
('D3 LNP', 0.0)
('D29 P1', 0.0)
('D29 LNP', 0.0)
('D28 P3', 0.0)
('D28 P2', 0.0)
('D28 LNP', 0.0)
('D27 P2', 0.0)
('D27 P1', 0.0)

('D27 LNP', 0.0)
('D26 P3', 0.0)
('D26 P2', 0.0)
('D26 P1', 0.0)
('D26 LNP', 0.0)
('D25 P2', 0.0)
('D25 P1', 0.0)
('D25 LNP', 0.0)
('D24 P3', 0.0)
('D24 P1', 0.0)
('D23 P1', 0.0)
('D23 LNP', 0.0)
('D22 P2', 0.0)
('D22 LNP', 0.0)
('D21 P3', 0.0)
('D21 LNP', 0.0)
('D20 P3', 0.0)
('D20 P2', 0.0)
('D20 LNP', 0.0)
('D2 P2', 0.0)
('D19 P3', 0.0)
('D19 P1', 0.0)
('D18 P2', 0.0)
('D18 LNP', 0.0)
('D17 P3', 0.0)
('D17 P2', 0.0)
('D17 LNP', 0.0)
('D16 P2', 0.0)
('D16 LNP', 0.0)
('D15 P3', 0.0)
('D15 P1', 0.0)
('D14 P1', 0.0)
('D14 LNP', 0.0)
('D13 P3', 0.0)
('D13 P2', 0.0)
('D12 LNP', 0.0)
('D11 P3', 0.0)
('D11 P2', 0.0)
('D11 LNP', 0.0)
('D10 P3', 0.0)
('D10 LNP', 0.0)
('D1 P2', 0.0)
('D1 LNP', 0.0)
('C9 P3', 0.0)
('C8 P3', 0.0)
('C8 P2', 0.0)
('C8 P1', 0.0)
('C8 LNP', 0.0)
('C7 P1', 0.0)
('C6 P3', 0.0)
('C6 LNP', 0.0)
('C5 P3', 0.0)
('C5 P2', 0.0)
('C4 P2', 0.0)
('C32 P2', 0.0)
('C32 P1', 0.0)
('C31 P2', 0.0)
('C31 P1', 0.0)
('C30 P2', 0.0)
('C3 P1', 0.0)
('C3 LNP', 0.0)
('C29 P1', 0.0)
('C29 LNP', 0.0)
('C28 P1', 0.0)
('C28 LNP', 0.0)
('C27 P3', 0.0)
('C27 P1', 0.0)
('C26 P3', 0.0)
('C26 P2', 0.0)
('C26 P1', 0.0)
('C26 LNP', 0.0)
('C25 P3', 0.0)
('C24 P3', 0.0)
('C24 P2', 0.0)
('C23 P3', 0.0)
('C22 P2', 0.0)
('C22 LNP', 0.0)
('C21 P3', 0.0)
('C21 P2', 0.0)
('C21 LNP', 0.0)
('C20 P3', 0.0)
('C20 P2', 0.0)
('C20 P1', 0.0)
('C2 P3', 0.0)
('C2 P2', 0.0)
('C2 P1', 0.0)
('C2 LNP', 0.0)
('C19 LNP', 0.0)
('C18 P3', 0.0)
('C18 P2', 0.0)
('C18 LNP', 0.0)
('C17 P2', 0.0)
('C16 P2', 0.0)
('C16 LNP', 0.0)
('C15 P3', 0.0)
('C15 P1', 0.0)
('C14 P1', 0.0)
('C14 LNP', 0.0)
('C13 P3', 0.0)
('C13 P2', 0.0)
('C13 P1', 0.0)
('C13 LNP', 0.0)
('C12 P3', 0.0)
('C12 P2', 0.0)
('C12 P1', 0.0)
('C11 LNP', 0.0)
('C10 P3', 0.0)
('C10 P2', 0.0)
('C10 P1', 0.0)
('C1 P2', 0.0)
('C1 P1', 0.0)
('B9 P3', 0.0)
('B9 LNP', 0.0)
('B8 P3', 0.0)
('B8 P2', 0.0)
('B8 LNP', 0.0)
('B7 P3', 0.0)
('B7 P2', 0.0)
('B7 LNP', 0.0)
('B6 P3', 0.0)
('B6 P2', 0.0)
('B6 P1', 0.0)
('B6 LNP', 0.0)
('B5 P3', 0.0)
('B5 LNP', 0.0)
('B4 P3', 0.0)
('B4 P2', 0.0)
('B4 P1', 0.0)
('B32 P3', 0.0)
('B32 P2', 0.0)
('B32 LNP', 0.0)
('B31 P1', 0.0)
('B31 LNP', 0.0)
('B30 P3', 0.0)
('B30 LNP', 0.0)
('B3 P3', 0.0)
('B3 LNP', 0.0)
('B29 P3', 0.0)
('B29 P2', 0.0)
('B29 P1', 0.0)
('B28 P3', 0.0)
('B28 P2', 0.0)
('B28 LNP', 0.0)
('B27 P2', 0.0)
('B27 P1', 0.0)
('B26 P3', 0.0)
('B26 P2', 0.0)

('B25 P1', 0.0)
('B24 P2', 0.0)
('B23 P3', 0.0)
('B23 P2', 0.0)
('B23 P1', 0.0)
('B23 LNP', 0.0)
('B22 P3', 0.0)
('B22 LNP', 0.0)
('B21 P3', 0.0)
('B21 P2', 0.0)
('B21 P1', 0.0)
('B21 LNP', 0.0)
('B20 P3', 0.0)
('B20 P2', 0.0)
('B20 P1', 0.0)
('B2 P3', 0.0)
('B2 P2', 0.0)
('B2 P1', 0.0)
('B2 LNP', 0.0)
('B19 P3', 0.0)
('B19 P2', 0.0)
('B19 P1', 0.0)
('B19 LNP', 0.0)
('B18 P3', 0.0)
('B17 P2', 0.0)
('B17 LNP', 0.0)
('B16 P3', 0.0)
('B16 P2', 0.0)
('B16 LNP', 0.0)
('B15 P3', 0.0)
('B15 P2', 0.0)
('B15 P1', 0.0)
('B14 P3', 0.0)
('B14 P2', 0.0)
('B14 P1', 0.0)
('B14 LNP', 0.0)
('B13 P3', 0.0)
('B13 P1', 0.0)
('B12 P3', 0.0)
('B12 P2', 0.0)
('B12 LNP', 0.0)
('B11 P2', 0.0)
('B10 P3', 0.0)
('B10 P2', 0.0)
('B10 LNP', 0.0)
('B1 P3', 0.0)
('B1 P2', 0.0)
('B1 P1', 0.0)
('B1 LNP', 0.0)
('A9 P3', 0.0)
('A9 P2', 0.0)
('A9 LNP', 0.0)
('A8 P3', 0.0)
('A8 P2', 0.0)
('A8 P1', 0.0)
('A8 LNP', 0.0)
('A7 P2', 0.0)
('A7 P1', 0.0)
('A6 P3', 0.0)
('A6 P2', 0.0)
('A6 P1', 0.0)
('A5 P2', 0.0)
('A5 P1', 0.0)
('A4 P2', 0.0)
('A4 P1', 0.0)
('A4 LNP', 0.0)
('A32 P3', 0.0)
('A32 P2', 0.0)
('A32 P1', 0.0)
('A32 LNP', 0.0)
('A31 P2', 0.0)
('A31 P1', 0.0)
('A3 P3', 0.0)
('A3 LNP', 0.0)
('A29 P3', 0.0)
('A28 P3', 0.0)
('A28 P2', 0.0)
('A28 P1', 0.0)
('A28 LNP', 0.0)
('A27 P3', 0.0)
('A27 P2', 0.0)
('A26 P3', 0.0)
('A26 P2', 0.0)
('A26 LNP', 0.0)
('A25 P3', 0.0)
('A25 P2', 0.0)
('A25 P1', 0.0)
('A24 P3', 0.0)
('A24 P2', 0.0)
('A24 LNP', 0.0)
('A23 P1', 0.0)
('A22 LNP', 0.0)
('A20 P3', 0.0)
('A20 P2', 0.0)
('A20 LNP', 0.0)
('A2 P2', 0.0)
('A2 LNP', 0.0)
('A19 P2', 0.0)
('A19 LNP', 0.0)
('A18 LNP', 0.0)
('A17 P3', 0.0)
('A17 P1', 0.0)
('A15 P3', 0.0)
('A15 P1', 0.0)
('A14 P3', 0.0)
('A14 P2', 0.0)
('A14 LNP', 0.0)
('A13 P3', 0.0)
('A13 P1', 0.0)
('A12 P3', 0.0)
('A12 P2', 0.0)
('A11 P3', 0.0)
('A11 P2', 0.0)
('A11 P1', 0.0)
('A1 P3', 0.0)
('A1 P1', 0.0)