GC33 Antibody Treatment against Liver Cancer Status Report 3

Melanie F. Pradier, Fernando Perez-Cruz

University Carlos III in Madrid

September 18th, 2015

Outline

Internship Overview: Key Questions

1. How well can we predict Survival?

2. Which features have an impact on Survival? How? $\rightarrow$ prognostic biomarkers (regardless of treatment).

3. Which features make the drug work? How? $\rightarrow$ predictive biomarkers.

   Notation: y: $N \times 1$ Time to First PD in months (Survival)
   X: $N \times D$ matrix of heterogeneous features

Focus: Survival Prediction

How well can we predict Survival?

- **Supervised approaches**

$$y = f(X) + \epsilon \tag{1}$$

- Prediction using 4 different methods:
  1. Linear Regression
  2. Lasso (Penalized LR)
  3. Gaussian Process
  4. Random Forest

- Identified Prognostic Variables according to Random Forest, at the top: CD56T, DN-PRE, CTROUGHC3D1-DD, NKP46-MESF, C Reactive Protein.

- Problems: too few data for supervised learning. All variables roughly same relevance level.

Focus: Biomarker Discovery
Which features have an impact on Survival? How?
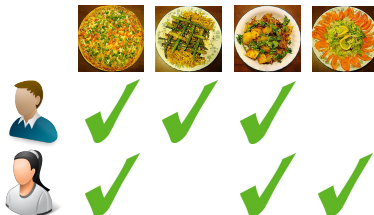
- **Unsupervised approaches**

$$f(y, X) \tag{2}$$

- We considered the following approaches:
  1. Dimensionality Reduction: Principal Component Analysis
  2. Dependency Metric: Map of Variables according to Mutual Information
  3. Latent Feature Model

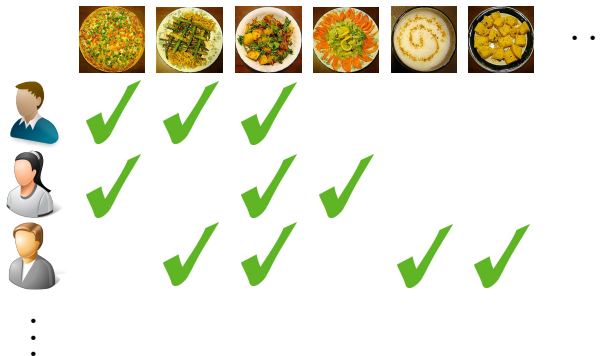- Focus: Biomarker Discovery using the Indian Buffet Process.

Overview
○○●○

Methodology

Joint Analysis: In the last meeting...

Two-step Analysis: first Placebo and then Drug
○○○○○○○○○○○○○○○○

Indian Buffet Process

Indian Buffet Process

Indian Buffet Process

Indian Buffet Process

Biomarker Discovery with the Indian Buffet Process

- Distribution over binary matrices $Z_{N \times K}$ where $K \to \infty$

- Finite N implies finite number of non-zero columns K_+.

$x \sim \mathcal{N}(\mu_0, \sigma_0^2)$

$Z \sim \mathrm{IBP}(\alpha)$
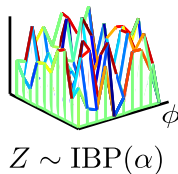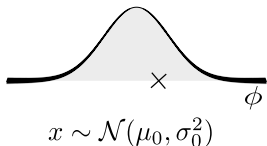
N patients
D covariates
K latent features

$$
\underset{X}{\underbrace{\boxed{}}^{N \times D}} = f \left(\underset{Z}{\underbrace{\boxed{}}^{D \times K}} \cdot \underset{B}{\underbrace{\boxed{}}^{K \times D}} \right)
$$

Outline

Methodology

1. Run IBP to identify interesting patient groups

2. Define group of interest G^* and reference G^B

3. Do Bootstrapping (to deal with low N)

4. Compute measure of Effect Size

5. Compute measure of Significance

Bootstrapping

- n_j: number of patients in group G_j

- for $l = 1 \cdots L$, do
    - Build new group $\widetilde{G_{jl}}$ by sampling n_j patients from G_j with replacement.
    - Compute statistic of interest.

- We then get a distribution of the statistic
  (more accurate $+$ uncertainty measure).

Measure of Effect Size

- For continuous variable d:

$$\beta_d = \frac{1}{L} \sum_{l=1}^{L} \log_2 \left(\frac{\mu_d(\widetilde{G_l^*})}{\mu_d(\widetilde{G_l^B})} \right) \tag{3}$$

- For categorical variable r:

$$\beta_r = \frac{1}{L} \sum_{l=1}^{L} \left(\mu_d(\widetilde{G_l^*}) - \mu_d(\widetilde{G_l^B}) \right) \tag{4}$$

Measure of Significance
Continuous Variables

For continuous variables, compute:

- Deviation compared to G^* variance

$$\gamma^* = \frac{\left|\mu_d(G^*) - \mu_d(G^B)\right|}{\sigma_d(G^*)} \qquad (5)$$

- Deviation compared to G^B variance

$$\gamma^B = \frac{\left|\mu_d(G^*) - \mu_d(G^B)\right|}{\sigma_d(G^B)} \qquad (6)$$

- T-test: Standard statistical test to compare two groups of data.

Measure of Significance
Categorical Variables

For categorical variables, compute:

- Distance to Binomial Mean
  - Fit a Binomial distribution to G^B
  - A variable r is considered significant if $\mu_r(G^*)$ is outside confidence interval

- Fisher Exact Test: Standard statistical test for contingency tables.

Outline

In the last meeting... 1/5

- Analysis of all patients (both Placebo and Drug)
- Placebo patients forced to signature: 1 0 0 0 0
- Identified group G^* of 25 patients with longer survival
- We defined 4 different reference groups G_j^B

```
1.1. pattern: 1  0  0  0  0      60 patients PLACEBO    survival: mean=3.22 median=1.52
1.2. pattern: 1  0  0  0  0      43 patients DRUG       survival: mean=2.80 median=2.04
 1. pattern: 1  0  0  0  0     103 patients            survival: mean=3.04 median=1.64
 2. pattern: 1  0  0  0  1       5 patients            survival: mean=1.54 median=1.22
 3. pattern: 1  0  0  1  0      15 patients            survival: mean=4.08 median=3.81
 4. pattern: 1  0  0  1  1       1 patients            survival: mean=4.07 median=4.07
 5. pattern: 1  0  1  0  0      20 patients            survival: mean=2.98 median=1.29
 6. pattern: 1  0  1  0  1       2 patients            survival: mean=1.22 median=1.22
 7. pattern: 1  0  1  1  0       8 patients            survival: mean=3.60 median=3.90
 8. pattern: 1  0  1  1  1       2 patients            survival: mean=10.83 median=10.82
 9. pattern: 1  1  0  0  0       3 patients            survival: mean=3.16 median=2.43
10. pattern: 1  1  0  1  0      10 patients            survival: mean=5.00 median=5.32
11. pattern: 1  1  0  1  1       1 patients            survival: mean=0.99 median=0.99
12. pattern: 1  1  1  0  0       5 patients            survival: mean=3.54 median=1.48
13. pattern: 1  1  1  0  1       2 patients            survival: mean=8.10 median=8.10
14. pattern: 1  1  1  1  0       2 patients            survival: mean=6.36 median=6.36
15. pattern: 1  1  1  1  1       1 patients            survival: mean=11.66 median=11.66
```

In the last meeting... 2/5

- Measure of Effect Size for Continuous Variables

In the last meeting... 3/5

- Measure of Significance for Continuous Variables

In the last meeting... 4/5
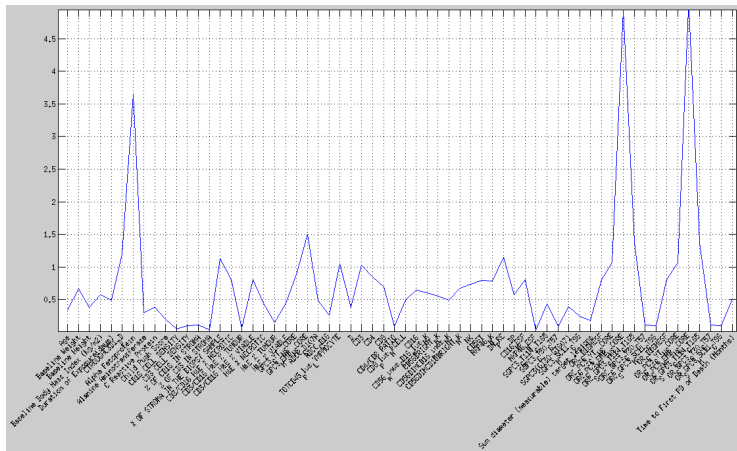
- Measure of Effect Size and Significance for Discrete Variables
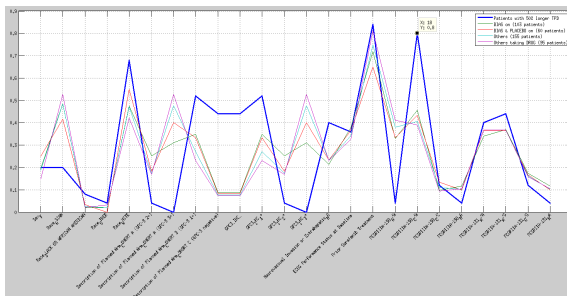
In the last meeting... 5/5

- Possible key variables are the following:
  - Lower GPC3-MEMB-H-SCORE
  - Lower Alpha Fetoprotein
  - Lower SGPC3-GT114-GT165
  - Higher probability for FCGRIIIA-158, C/A (lower allele A/A)
  - Higher DN-PRE
  - Higher probability for GPC3-IHC- (lower GPC3-IHC-3+)
  - ...

- Some problems:
  - Unexpected direction for variables in blue
  - Prognostic + Predictive variables mixed

Which variables impact survival Vs which ones make the treatment work?

Outline

First Placebo

- Run IBP on Placebo patients

```
nk = 60   20    9
ALL                                 meanTPFD=3.22, medianTFPD=1.52
1. Pattern: 1  0  0 numPat=33, meanTPFD=3.17, medianTFPD=1.48
2. Pattern: 1  0  1 numPat= 7, meanTPFD=4.51, medianTFPD=3.22
3. Pattern: 1  1  0 numPat=18, meanTPFD=2.43, medianTFPD=1.46
4. Pattern: 1  1  1 numPat= 2, meanTPFD=6.55, medianTFPD=6.55
```

- Identify Interesting groups
  - G_{strong}^{P}: Placebo patients with longer survival naturally (patterns 2,4)
  - G_{normal}^{P}: Placebo patients having an average survival
    - G_{B1}^{P}: pattern 1
    - G_{B2}^{P}: pattern 3
    - G_{B3}^{P}: pattern 1 & 3, complementary to G_{strong}^{P}
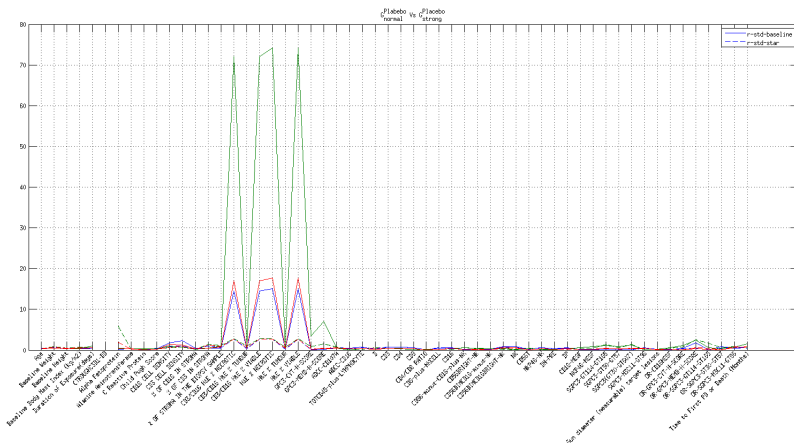
First Placebo
Measure of Effect Size
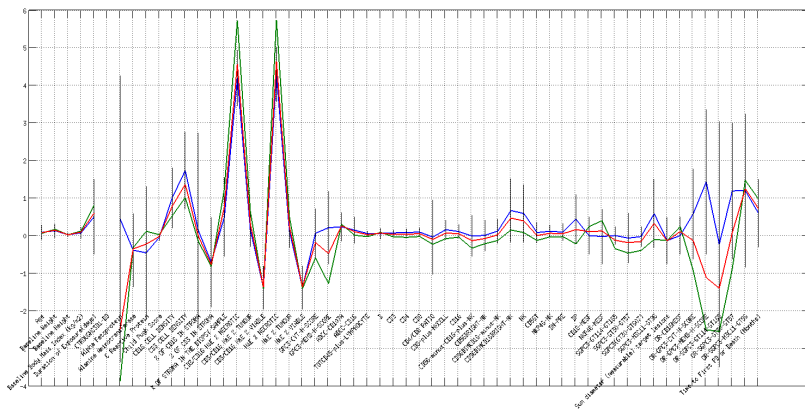
- Measure of Effect Size for Continuous Variables

First Placebo
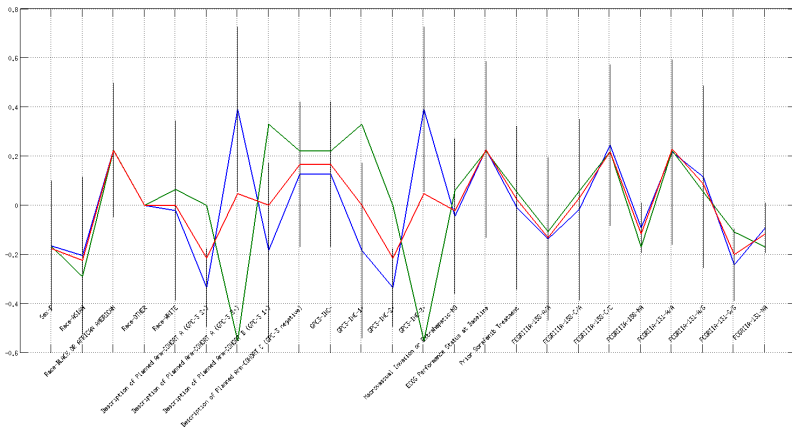Measure of Significance for continuous variables

- Measure of Significance for continuous variables

First Placebo
Measure of Effect Size for discrete variables

- Measure of Effect Size for discrete variables

First Placebo
Measure of Significance for continuous/discrete variables

- Statistical measures of Effect Significance


```
T-TEST
              H&E % NECROTIC:4.6818e-17
               H&E % VIABLE:4.6818e-17
      CD3/CD16 H&E % NECROTIC:1.0417e-15
       CD3/CD16 H&E % VIABLE:1.0417e-15
              CD3 CELL DENSITY:7.8890e-04
             CD16 CELL DENSITY:4.6483e-03
      OR-GPC3-MEMB-H-SCORE:3.3821e-02
```

```
FISHER TEST

              FCGRIIA-131-G/G:1.0825e-05
                   Sex-F:1.0401e-04
          FCGRIIIA-158-NA:1.9980e-04
           FCGRIIA-131-NA:1.9980e-04
              Race-ASIAN:2.1021e-04
        FCGRIIIA-158-A/A:7.0105e-04
        FCGRIIIA-158-C/A:2.9941e-02
```

First Placebo
Conclusions

- Two communities in Placebo
- Patients have longer survival if (high impact):
    - High CD3/CD16 H&E % NECROTIC
    - Low CD3/CD16 H&E % VIABLE
    - High H&E % NECROTIC
    - Low H&E % VIABLE

- Smaller Effect Size:
    - Lower probability for FCGRIIIA-131, G/G
    - Higher probability for FCGRIIIA-158, C/A (lower allele A/A)
    - Lower probability for Asian
    - Lower probability for Female

Second Drug

- Fix latent features for Placebo patients and run IBP with all patients.
- We only infer features of Drug patients

```
nk =  120    44    19    28
DRUG                              meanTPFD=3.54, medianTPFD=2.62
 1. Pattern: 1  0  0  0 numPat=46, meanTPFD=3.40, medianTPFD=2.83
 2. Pattern: 1  0  0  1 numPat=17, meanTPFD=3.83, medianTPFD=3.81
 3. Pattern: 1  0  1  0 numPat=10, meanTPFD=2.46, medianTPFD=1.30
 4. Pattern: 1  0  1  1 numPat= 3, meanTPFD=2.10, medianTPFD=1.25
 5. Pattern: 1  1  0  0 numPat=32, meanTPFD=3.30, medianTPFD=1.38
 6. Pattern: 1  1  0  1 numPat= 6, meanTPFD=4.10, medianTPFD=3.93
 7. Pattern: 1  1  1  0 numPat= 4, meanTPFD=5.12, medianTPFD=2.08
 8. Pattern: 1  1  1  1 numPat= 2, meanTPFD=10.83, medianTPFD=10.82
```

- One new feature created
- When 4th feature is active, generally longer survival

Second Drug
Identification of Interesting Groups

- Identify Interesting groups: $G_{normal}^{Placebo}$, G_{normal}^{Drug-}, G_{normal}^{Drug+}, $G_{strong}^{Placebo}$, G_{strong}^{Drug-}, G_{strong}^{Drug+}
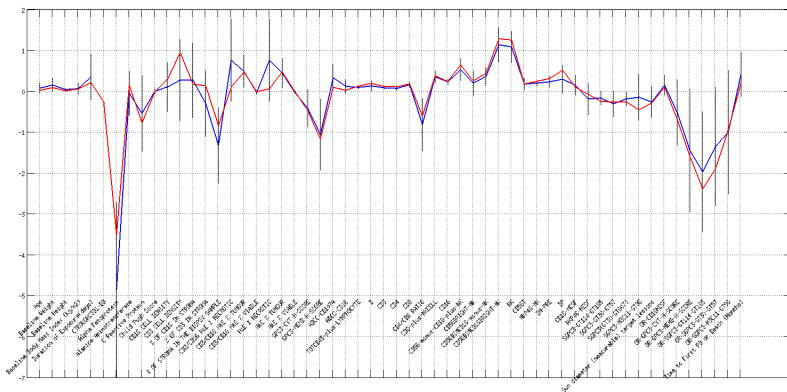
Patients Categorization: X P Mean Median (TFPD)		
	Placebo	Drug $-$	Drug $+$
Normal (total)	**51P 2.91 1.47**	**78P 3.36 2.24**	**23P 3.90 3.84**
$100-$	33P 3.17 1.48	46P 3.40 2.83	17P 3.83 3.81
$110-$	18P 2.43 1.46	32P 3.30 1.38	06P 4.10 3.93
Strong (total)	**09P 4.96 3.96**	**14P 3.22 1.52**	**05P 5.59 5.08**
$101-$	07P 4.51 3.22	10P 2.46 1.30	03P 2.10 1.25
$111-$	02P 6.55 6.55	04P 5.12 2.08	02P 10.83 10.82
Last Feature	-0	-0	-1

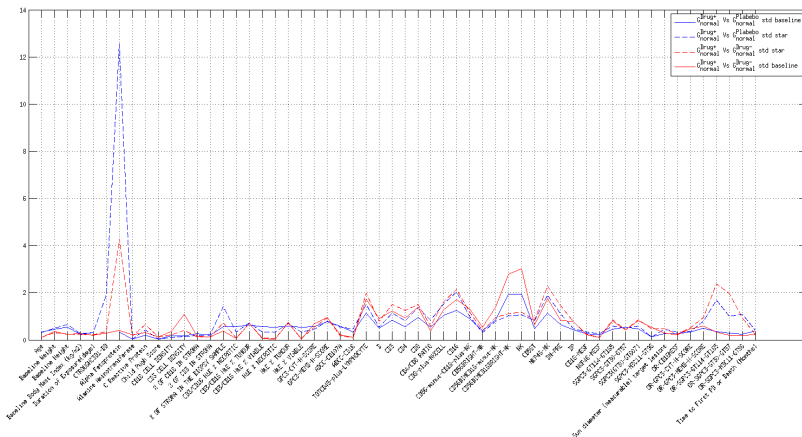# Second Drug
Measure of Effect Size

- Measure of Effect Size for Continuous Variables

Second Drug
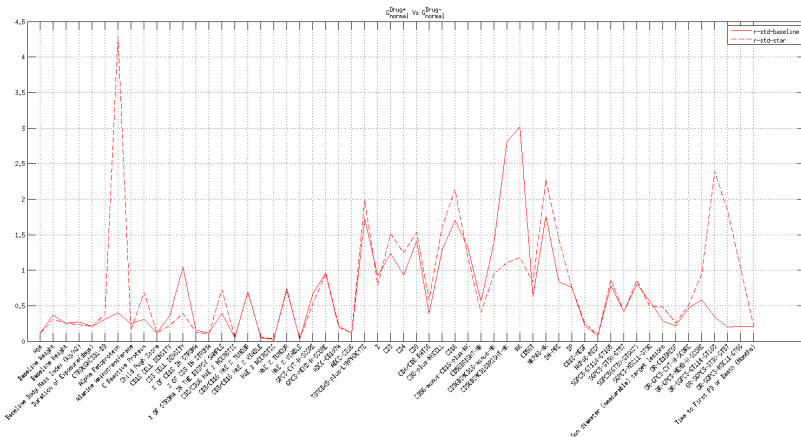Measure of Significance for continuous variables

- Measure of Significance for continuous variables

Second Drug
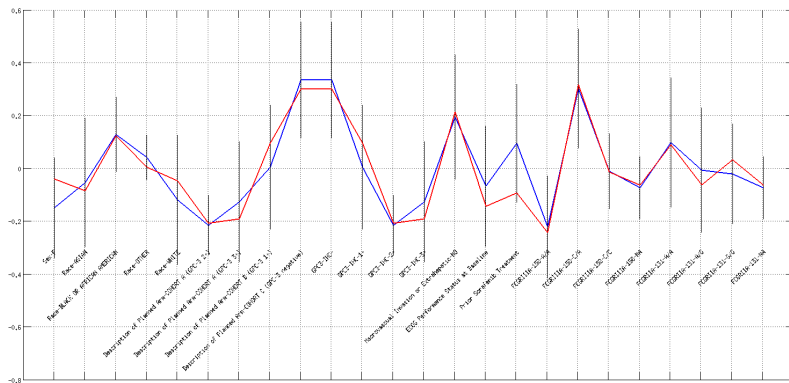## Measure of Significance for continuous variables

- Measure of Significance for continuous variables

Second Drug
Measure of Effect Size for discrete variables

- Measure of Effect Size for discrete variables

Second Drug
Measure of Significance for discrete variables

- Measure of Effect Significance for discrete variables

Overview Methodology Joint Analysis: In the last meeting... **Two-step Analysis: first Placebo and then Drug**
○○○○ ○○○○○○○○○○○○○○●○

Second Drug

Second Drug

- Possible predictive variables are the following:
    - Lower Alpha Fetoprotein
    - Lower OR-SGPC3-GT114-GT165
    - Higher NK
    - Higher CD56DIMCD16BRIGHT-NK

    - Higher probability for FCGRIIIA-158, C/A
    - Higher probability for GPC3-IHC- (lower GPC3-IHC-2+, GPC3-IHC-3+)

Conclusion

- From First Placebo Analysis, prognostic variables:
  - High CD3/CD16 H&E % NECROTIC
  - Low CD3/CD16 H&E % VIABLE
  - High H&E % NECROTIC
  - Low H&E % VIABLE
  - Lower probability for FCGRIIIA-131, G/G
  - Higher probability for FCGRIIIA-158, C/A (lower allele A/A)

- From Second Drug Analysis, predictive variables:
  - Lower Alpha Fetoprotein
  - Lower OR-SGPC3-GT114-GT165
  - Higher NK
  - Higher CD56DIMCD16BRIGHT-NK
  - Higher probability for FCGRIIIA-158, C/A
  - Higher probability for GPC3-IHC- (lower GPC3-IHC-2+,
    GPC3-IHC-3+)