Bayesian modeling for biomarker discovery in clinical trials

Melanie F. Pradier

University Carlos III in Madrid

Collaborators: F. Milletti, O. Puig at Roche Innovation Center, New York and F. Perez-Cruz at Bell Labs, New York

May 21th, 2016

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Biomarkers are used everywhere!!

Some examples

- Prostate-specific antigen (PSA) to diagnose prostate cancer
- Estrogen / progesterone to predict sensitivity to endocrine therapy in breast cancer
- KRAS mutation to predict resistance to EGFr antibody treatment

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Biomarkers are used everywhere!!

Some examples

- Prostate-specific antigen (PSA) to diagnose prostate cancer
- Estrogen / progesterone to predict sensitivity to endocrine therapy in breast cancer
- KRAS mutation to predict resistance to EGFr antibody treatment

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

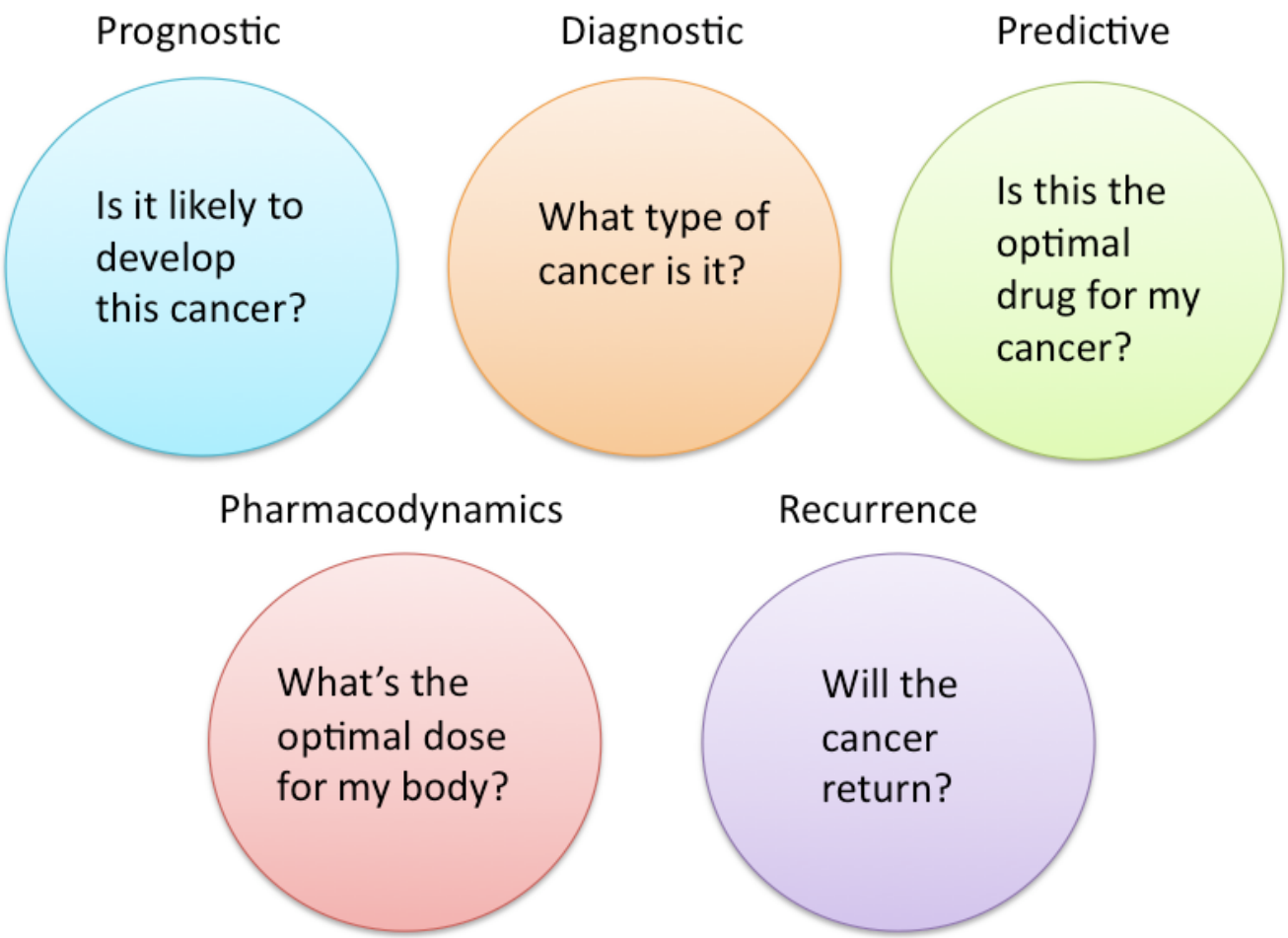
Biomarkers are used everywhere!!

Some examples

- Prostate-specific antigen (PSA) to diagnose prostate cancer
- Estrogen / progesterone to predict sensitivity to endocrine therapy in breast cancer
- KRAS mutation to predict resistance to EGFr antibody treatment

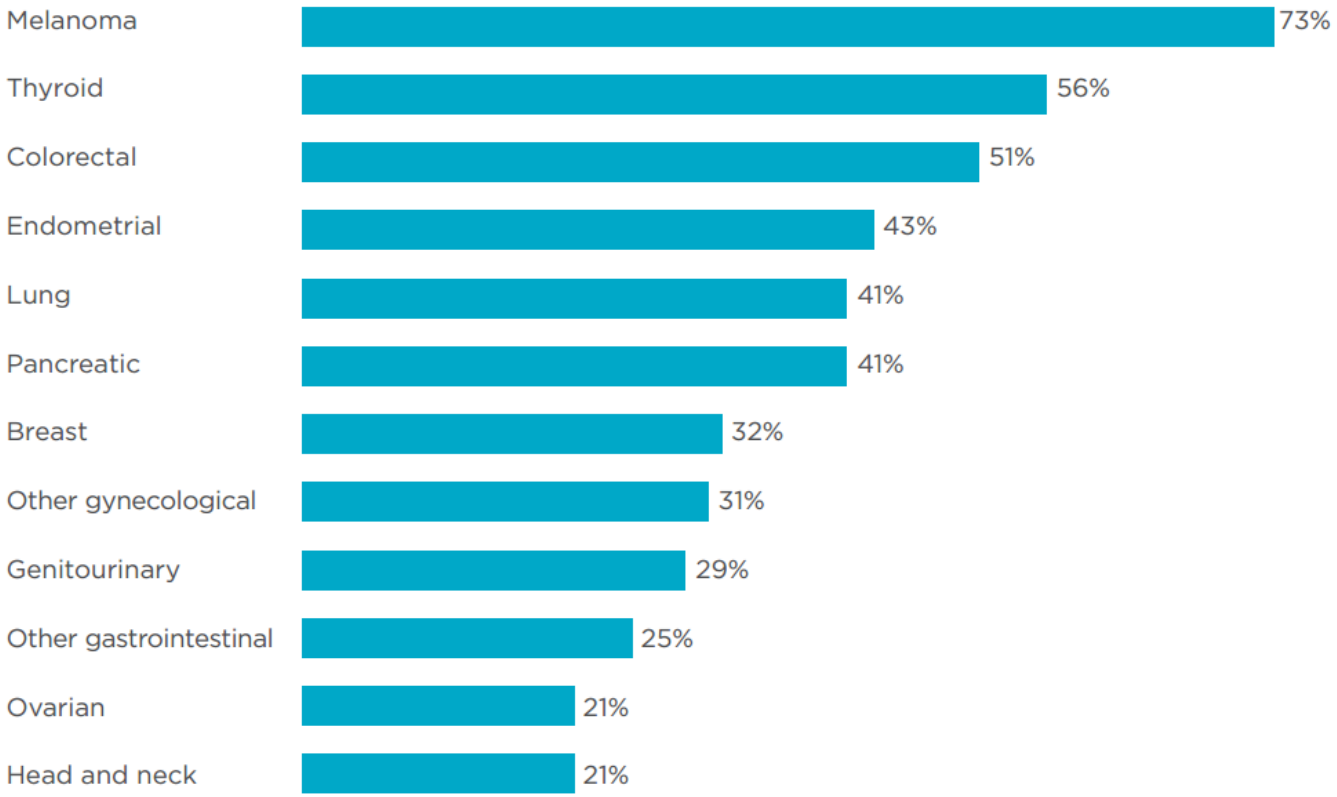
Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".



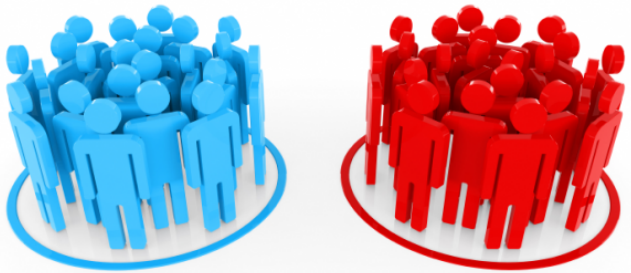
Biomarkers as potential targets for new drugs

TACKLING TUMORS: Percentage of patients whose tumors were driven by certain genetic mutations that could be targets for specific drugs, by types of cancer.



Source: *Wall Street Journal* Copyright 2011 by DOW JONES & COMPANY, INC. Reproduced with permission of DOW JONES & COMPANY, INC.

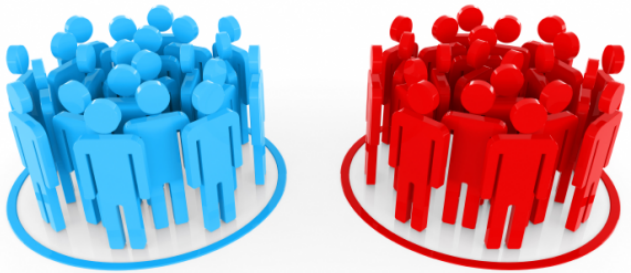
Clinical Trial Scenario



We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers
3. Biomarkers as potential new targets for drugs

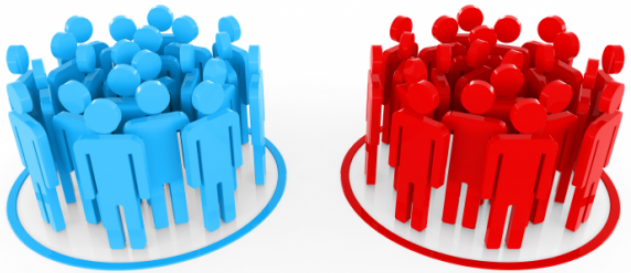
Clinical Trial Scenario



We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers
3. Biomarkers as potential new targets for drugs

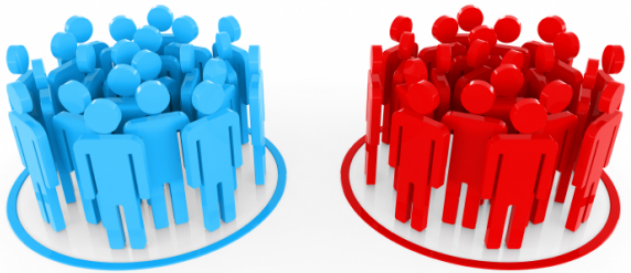
Clinical Trial Scenario



We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers
3. Biomarkers as potential new targets for drugs

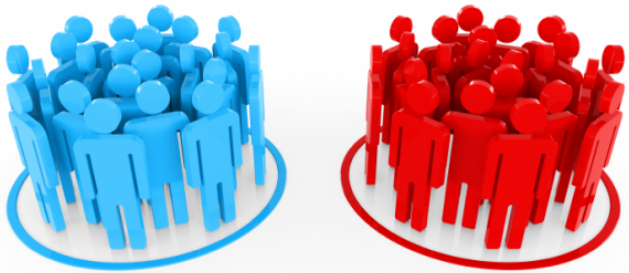
Clinical Trial Scenario



We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers
3. Biomarkers as potential new targets for drugs

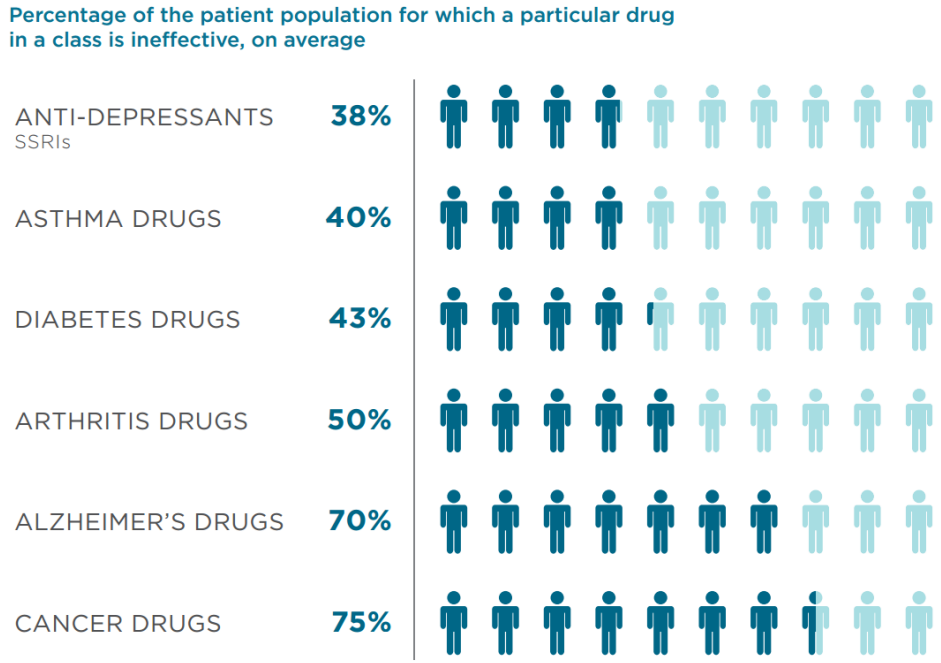
Clinical Trial Scenario



We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers
3. Biomarkers as potential new targets for drugs

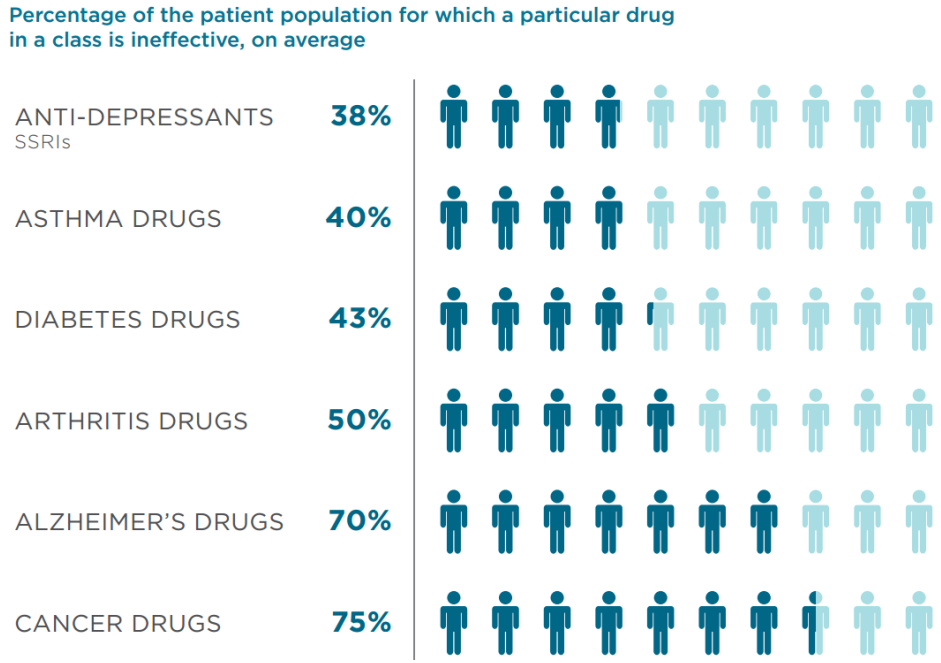
Population Heterogeneity



Responses are practically unique! (Wilkinson et.al., 2005)

- most major drugs only work for 25 to 65% of patients.
- 2M cases of adverse drug reactions, including 100.000 deaths.

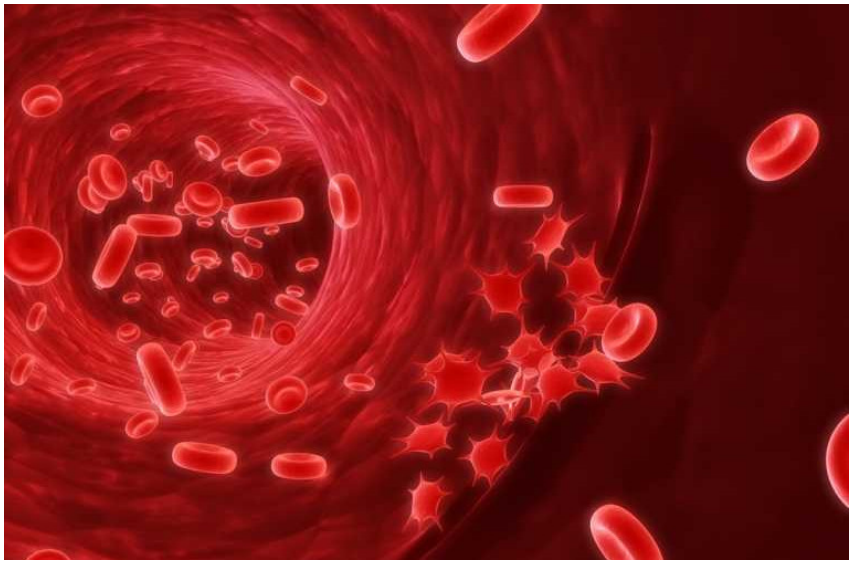
Population Heterogeneity



Responses are practically unique! (Wilkinson et.al., 2005)

- most major drugs only work for 25 to 65% of patients.
- 2M cases of adverse drug reactions, including 100.000 deaths.

And as a consequence...

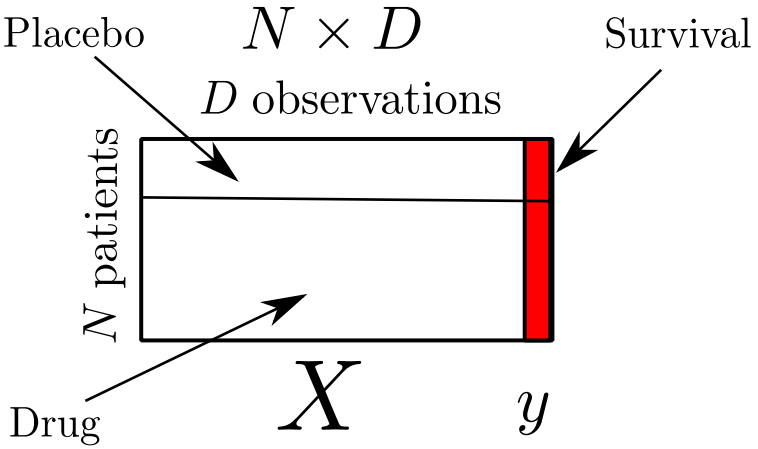


It is hard to isolate drug effect from natural response

- Sometimes you can (ex: cross-over designs)
- But in general, still hard to do

Our approach
A probabilistic perspective

- Focus on latent feature models.

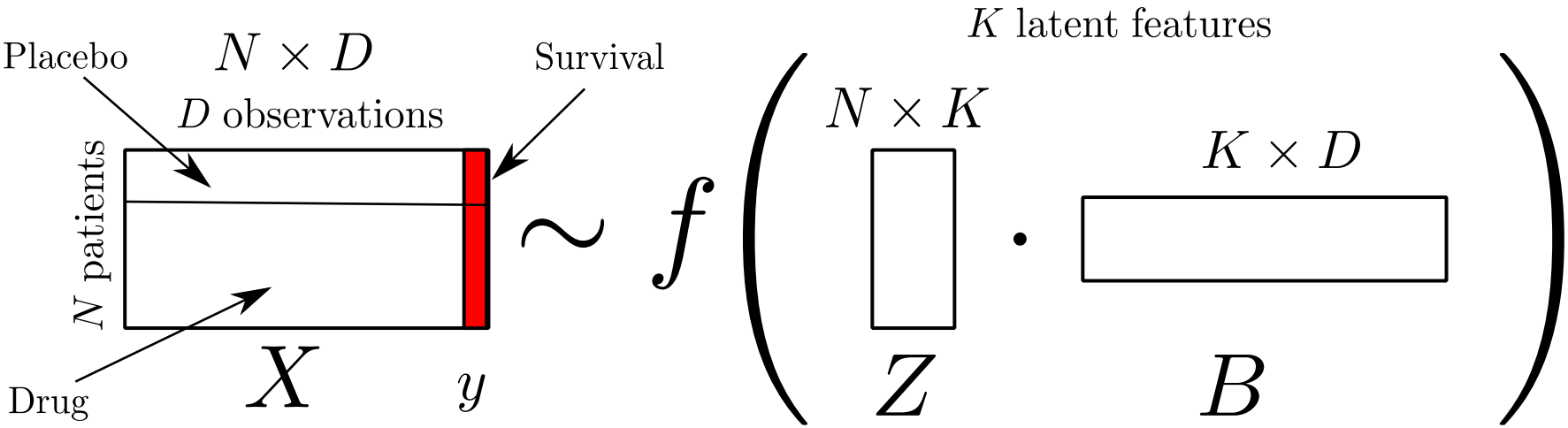


$$Z \sim \text{Indian Buffet Process}(\alpha) \quad (1)$$

- Bayesian: Put a prior over assumptions
- Non-parametric: Model complexity, i.e., number of latent features, is also inferred

Our approach
A probabilistic perspective

- Focus on latent feature models.



$$Z \sim \text{Indian Buffet Process}(\alpha) \qquad (1)$$

- Bayesian: Put a prior over assumptions
- Non-parametric: Model complexity, i.e., number of latent features, is also inferred

Sub-populations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug.

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean TFPD (months)	Median TFPD (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Sub-populations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug.

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean TFPD (months)	Median TFPD (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Sub-populations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug.

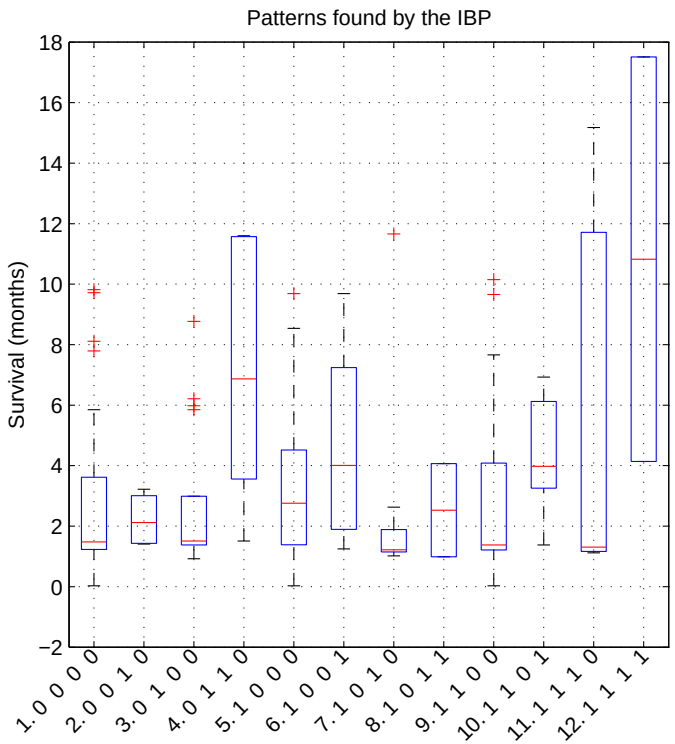
Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean TFPD (months)	Median TFPD (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Sub-populations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

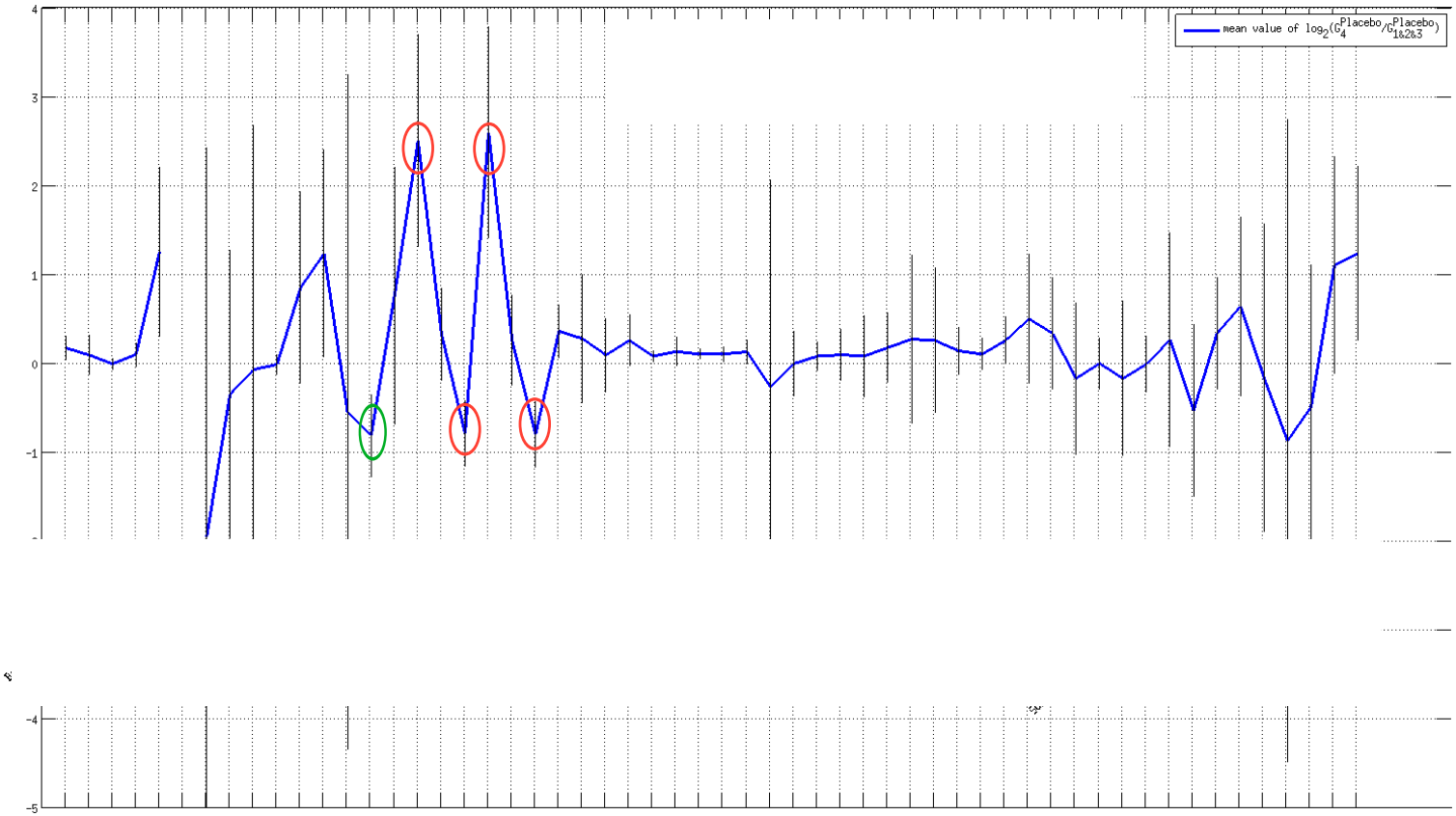
- 180 patients: 60 took a placebo, 120 took the drug.

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean TFPD (months)	Median TFPD (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01



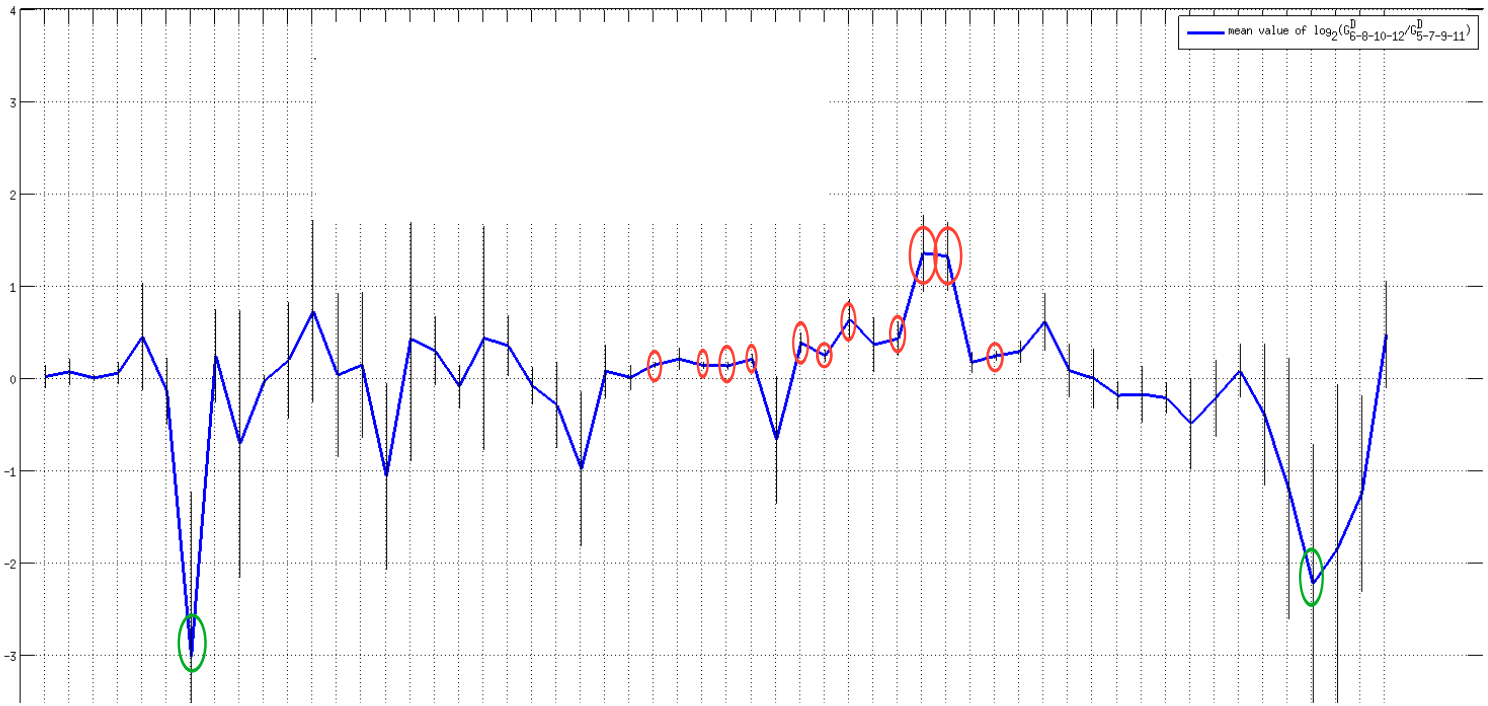
Prognostic biomarkers

Strong Placebo Vs Normal Placebo



Predictive biomarkers

Strong Drug Vs Normal Drug



Conclusions

Accounting for population heterogeneity is crucial!

With Bayesian models, we can...

1. Identify sub-populations
2. Identify prognostic and predictive biomarkers
3. Either globally or community specific

Conclusions

Accounting for population heterogeneity is crucial!

With Bayesian models, we can...

1. Identify sub-populations
2. Identify prognostic and predictive biomarkers
3. Either globally or community specific

Conclusions

Accounting for population heterogeneity is crucial!

With Bayesian models, we can...

1. Identify sub-populations
2. Identify prognostic and predictive biomarkers
3. Either globally or community specific

THE BAYESIAN REVOLUTION IN GENETICS

Mark A. Beaumont and Bruce Rannala‡*

Bayesian statistics allow scientists to easily incorporate prior knowledge into their data analysis. Nonetheless, the sheer amount of computational power that is required for Bayesian statistical analyses has previously limited their use in genetics. These computational constraints have now largely been overcome and the underlying advantages of Bayesian approaches are putting them at the forefront of genetic data analysis in an increasing number of areas.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Why Bayesian models?

Advantages of being Bayesian

1. Priors to constrain solution space
2. Uncertainty measure
3. Model averaging

"All models are wrong, but some are useful" - G. E. P. Box

Outlook

- Bottleneck: computational cost.
- Incorporate expert knowledge.

Acknowledgments

Current and future research results at: www.melaniefpradier.work

Email: melanie@tsc.uc3m.es

- Fernando Perez-Cruz
- Francesca Miletti
- Oscar Puig
- TSC at UC3M
- Marie-Curie ITN-MLPM

Thank you!