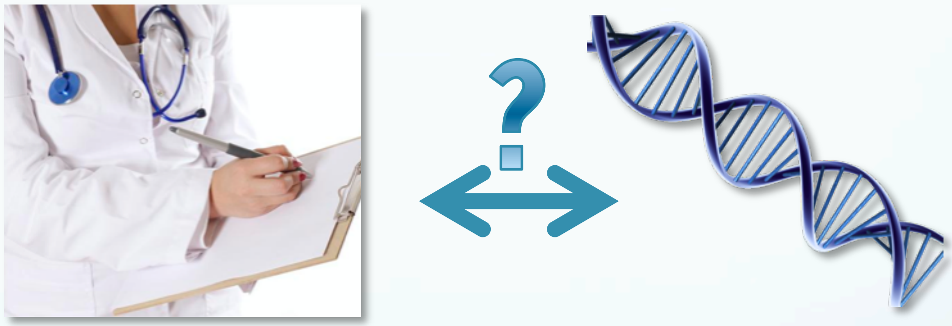
Probabilistic Analysis of Genetic Associations with Clinical Features in Cancer

Melanie F. Pradier, Julia E. Vogt, Stefan Stark, Theofanis Karaletsos, Fernando Perez-Cruz, Gunnar Rätsch

University Carlos III in Madrid & Memorial Sloan-Kettering Cancer Center

Motivation

- Objective: Find meaningful genotype-phenotype relationships.



Motivation

- Objective: Find meaningful genotype-phenotype relationships.

Why does it matter?

- Improved Diagnosis
- Risk Identification
- Biological Insight

Motivation

- Objective: Find meaningful genotype-phenotype relationships.

Why does it matter?

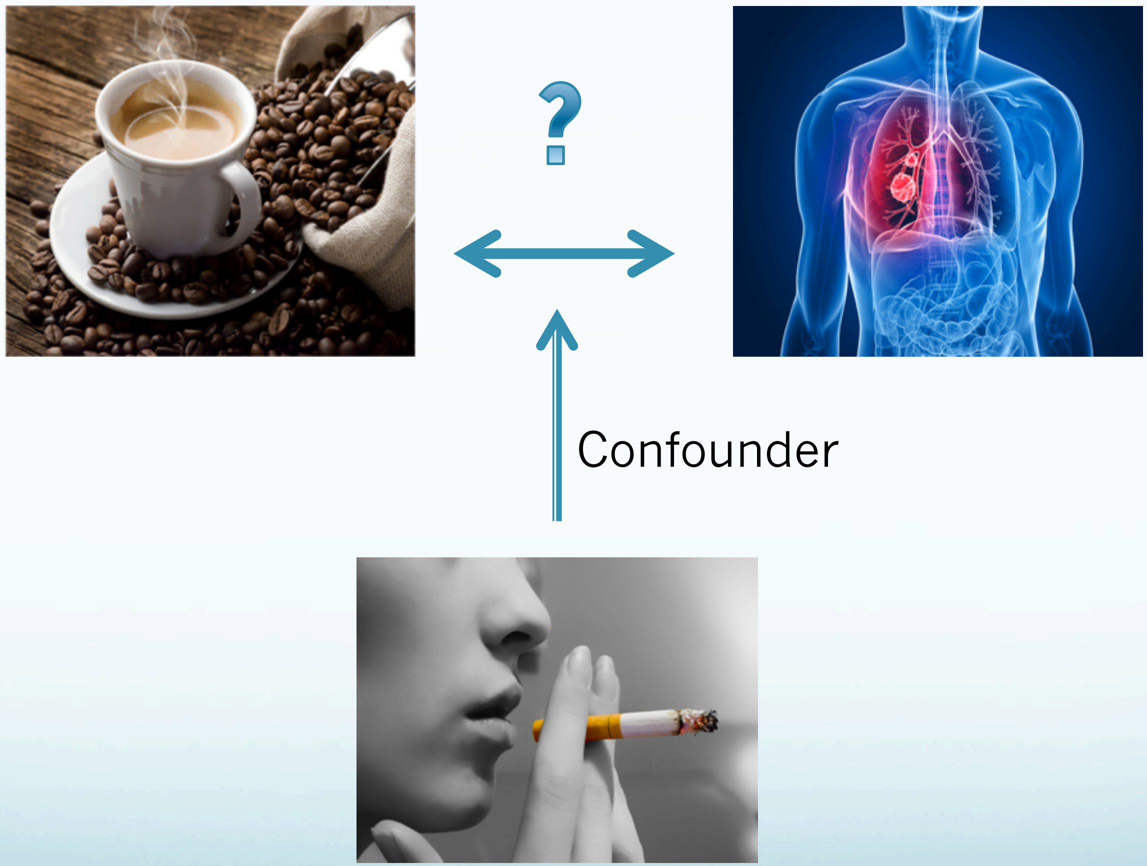
- Improved Diagnosis
- Risk Identification
- Biological Insight

We should be aware of…

We should be aware of…

?

We should be aware of…

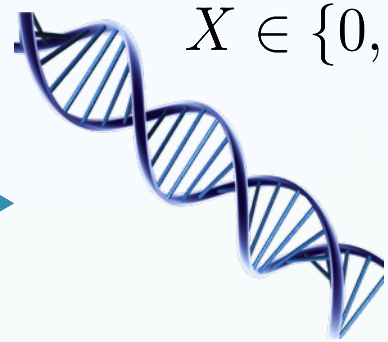


Problem Description

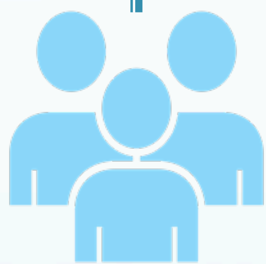
Problem Description

$Y \in \mathbb{N}^{P \times Q}$

$X \in \{0,1\}^{P \times G}$

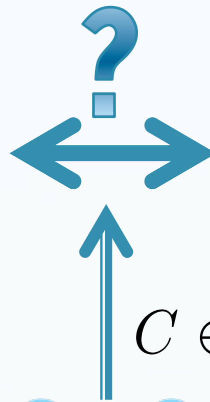


$C \in \{0,1\}^{P \times L}$

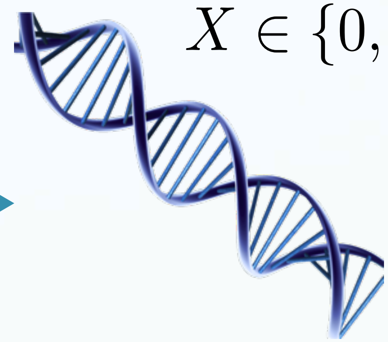


Problem Description

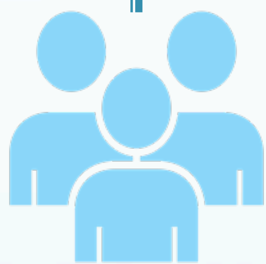
$Y \in \mathbb{N}^{P \times Q}$

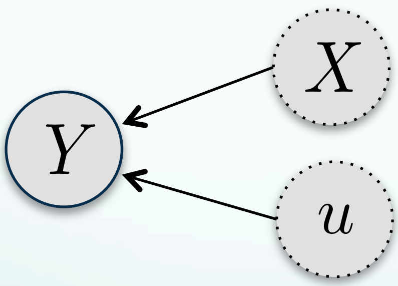


$X \in \{0,1\}^{P \times G}$



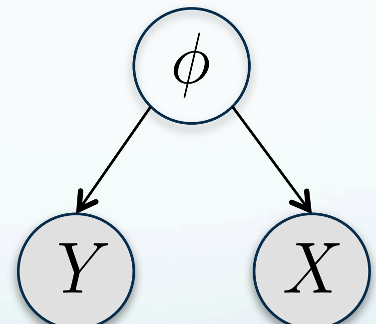
$C \in \{0,1\}^{P \times L}$





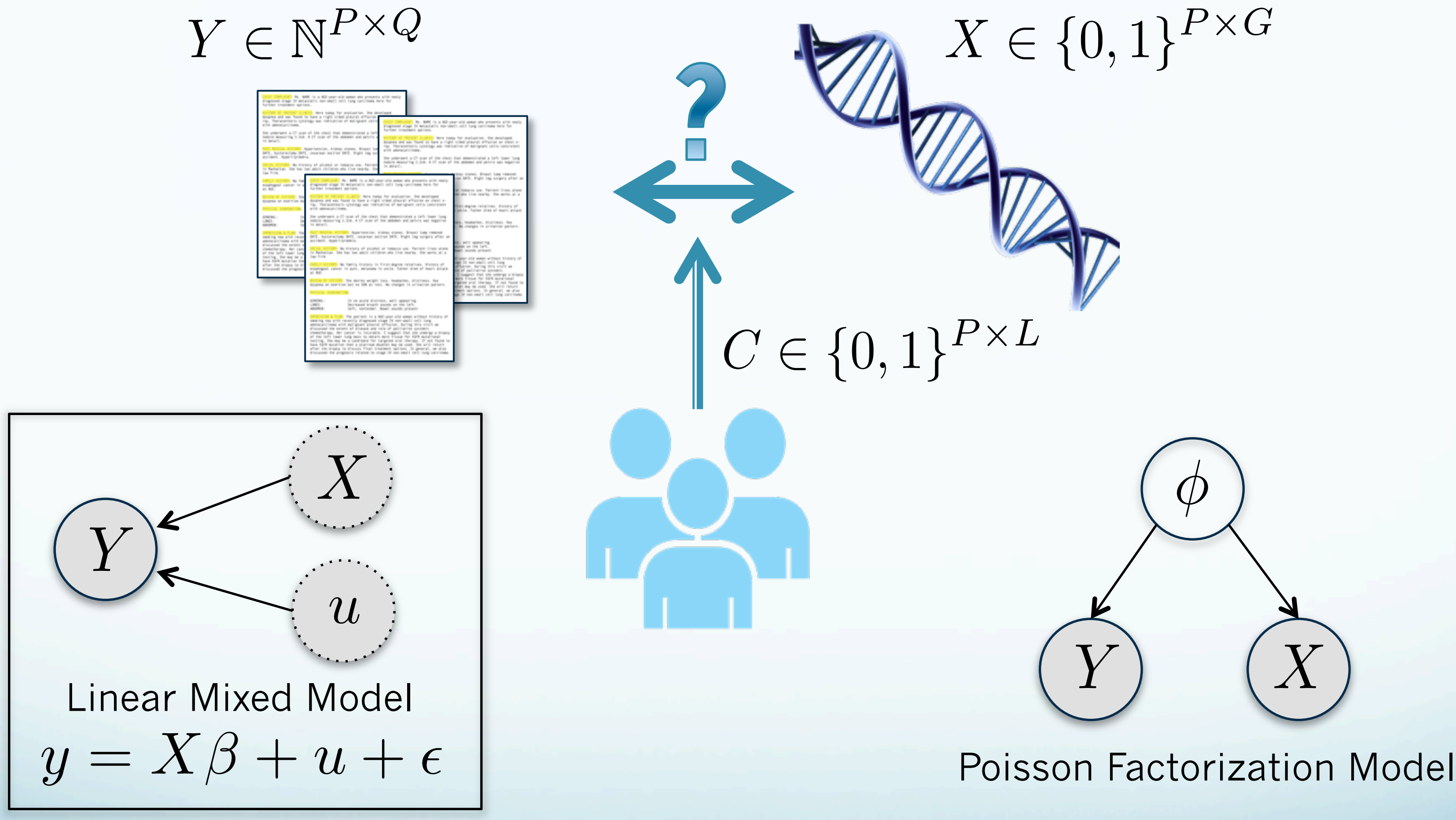
Linear Mixed Model

$$y = X\beta + u + \epsilon$$



Poisson Factorization Model

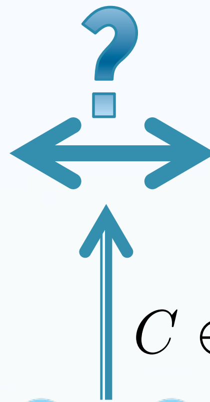
Problem Description

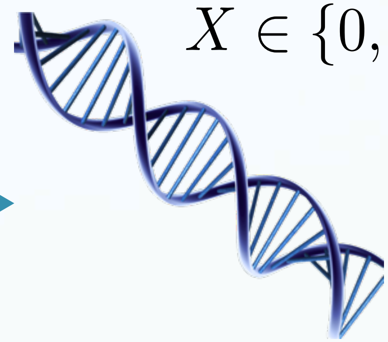


[1] C. Lippert, J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman, "FaST linear mixed models for genome-wide association studies," *Nat Meth*, vol. 8, no. 10, pp. 833–835, Oct. 2011.

Problem Description

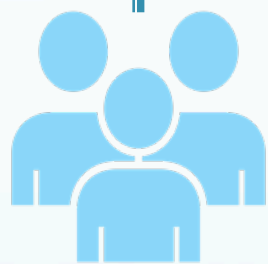
$Y \in \mathbb{N}^{P \times Q}$

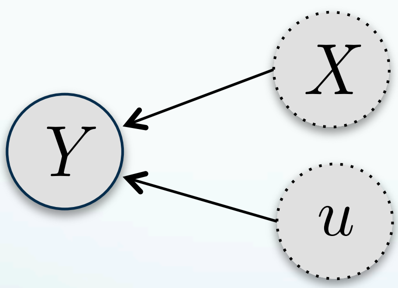




$X \in \{0,1\}^{P \times G}$

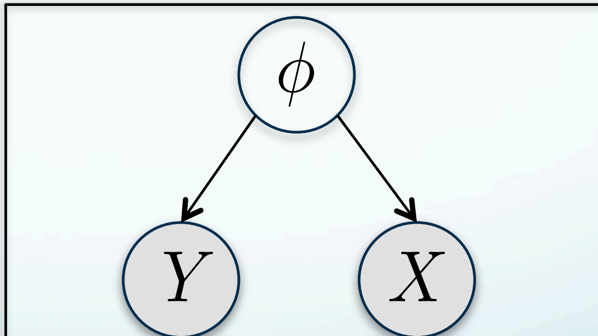
$C \in \{0,1\}^{P \times L}$





Linear Mixed Model

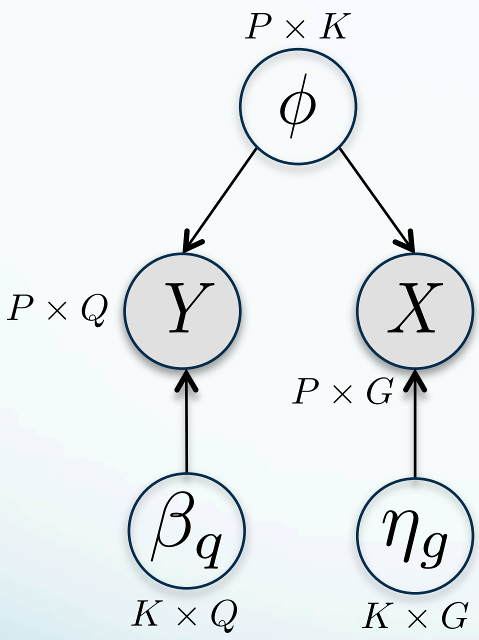
$$y = X\beta + u + \epsilon$$



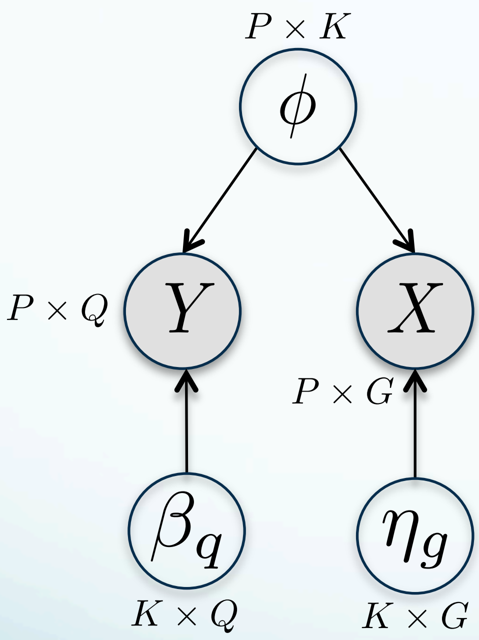
Poisson Factorization Model

[1] C. Lippert, J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman, "FaST linear mixed models for genome-wide association studies," *Nat Meth*, vol. 8, no. 10, pp. 833–835, Oct. 2011.
[2] P. Gopalan, D. Blei, "Content-based recommendations with Poisson factorization," presented at the Advances in Neural Information Processing Systems 27, 2014.

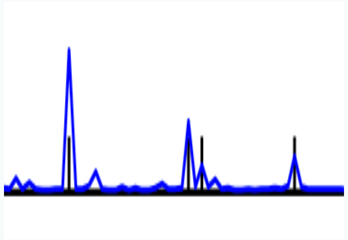
Poisson Factorization Model



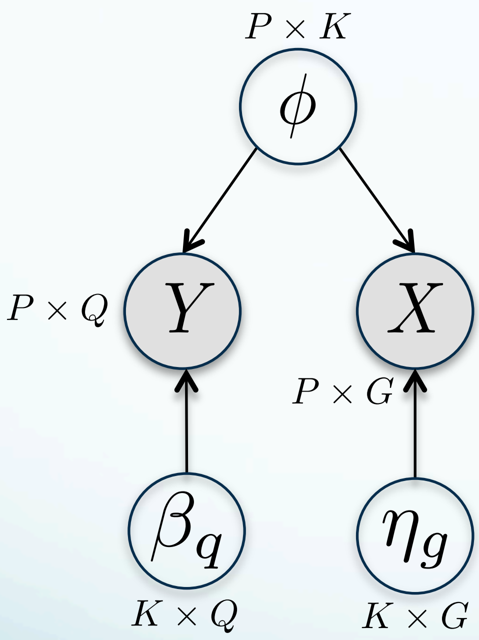
Poisson Factorization Model



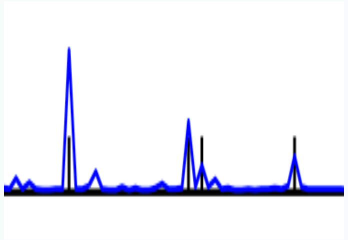
- Clinical factors β_q
- Genetic factors η_g



Poisson Factorization Model

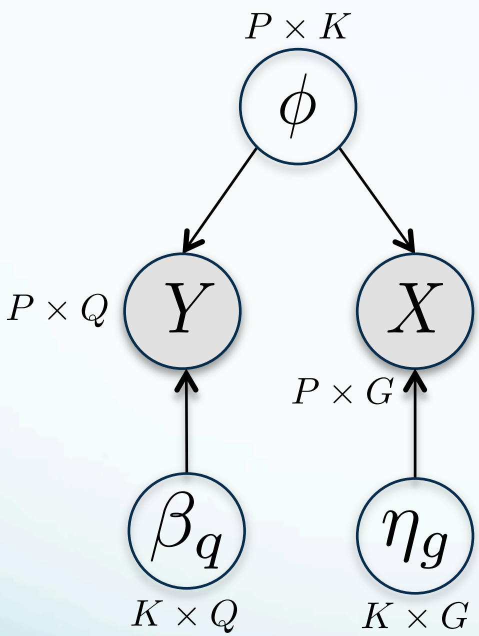


- Clinical factors β_q
- Genetic factors η_g

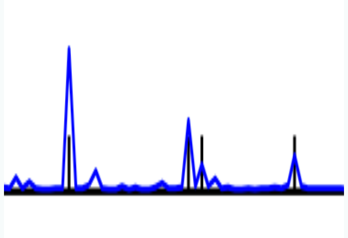


$$x_{pg} \sim \text{Poisson}(\phi_p \eta_g)$$
$$y_{pq} \sim \text{Poisson}(\phi_p \beta_q)$$
$$\text{others} \sim \text{Gamma}(a, b)$$

Poisson Factorization Model



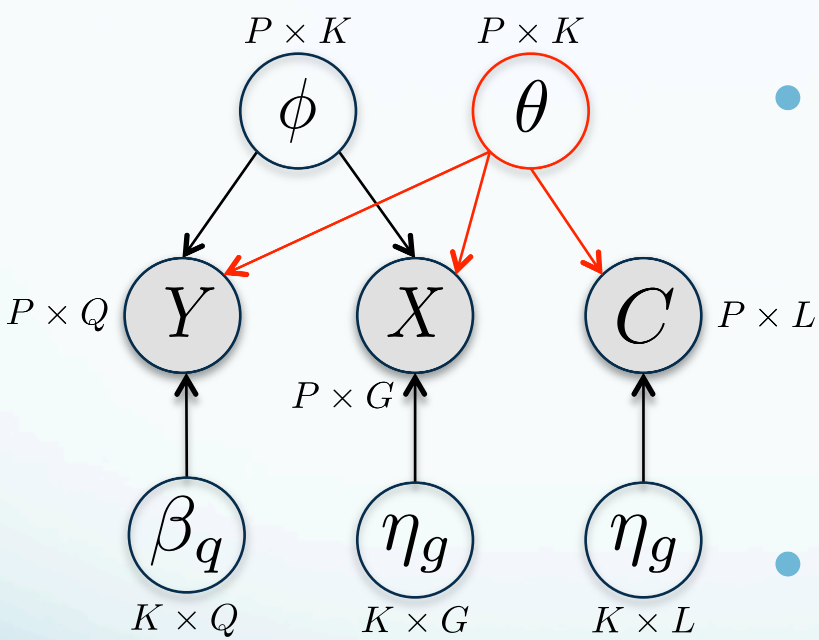
- Clinical factors β_q
- Genetic factors η_g



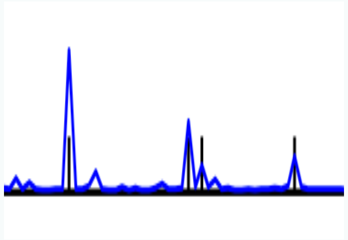
$$x_{pg} \sim \text{Poisson}(\phi_p \eta_g)$$
$$y_{pq} \sim \text{Poisson}(\phi_p \beta_q)$$
$$\text{others} \sim \text{Gamma}(a, b)$$

- ϕ captures structure

Poisson Factorization Model



- Clinical factors β_q
- Genetic factors η_g



$$x_{pg} \sim \mathrm{Poisson}(\phi_p \eta_g + \theta_p \eta'_g)$$
$$y_{pq} \sim \mathrm{Poisson}(\phi_p \beta_q + \theta_p \beta'_q)$$
$$\text{others} \sim \mathrm{Gamma}(a, b)$$

- ϕ captures interesting structure
- θ explains away confounders

Some Results

pelvic	0.033652
ovarian	0.031157
vaginal	0.022970
endometria	0.022705
woman	0.019036
recurrent	0.017424
absent	0.016078
ca	0.015917
female	0.014931
surgically	0.013436

TP53	0.026025
KRAS	0.023327
ARID1A	0.017346
PPP2R1A	0.013243
PIK3CA	0.011006

prostate	0.119886
psa	0.062841
gleason	0.030116
prostatectomy	0.018958
adenocarcinoma	0.017361
androgen	0.012359
protocol	0.011871
lupron	0.011870
urinary	0.011405
radical	0.011151

SPOP	**0.015479**
FOXA1	0.014104
SPEN	0.012706
MLL3	0.012247
AR	**0.012224**

colon	0.061547
folfox	0.032459
woman	0.021636
vemurafenib	0.017319
appreciated	0.014151
cea	0.013780
folfiri	0.013745
braf	0.013290
involving	0.011781
bevacizumab	0.010973

APC	**0.034585**
KRAS	**0.032768**
TP53	0.025132
BRAF	**0.023445**
SMAD4	**0.014745**

Some Results

pelvic	0.033652
ovarian	0.031157
vaginal	0.022970
endometria	0.022705
woman	0.019036
recurrent	0.017424
absent	0.016078
ca	0.015917
female	0.014931
surgically	0.013436

TP53	0.026025
KRAS	0.023327
ARID1A	0.017346
PPP2R1A	0.013243
PIK3CA	0.011006

prostate	0.119886
psa	0.062841
gleason	0.030116
prostatectomy	0.018958
adenocarcinoma	0.017361
androgen	0.012359
protocol	0.011871
lupron	0.011870
urinary	0.011405
radical	0.011151

SPOP	**0.015479**
FOXA1	0.014104
SPEN	0.012706
MLL3	0.012247
AR	**0.012224**

colon	0.061547
folfox	0.032459
woman	0.021636
vemurafenib	0.017319
appreciated	0.014151
cea	0.013780
folfiri	0.013745
braf	0.013290
involving	0.011781
bevacizumab	0.010973

APC	**0.034585**
KRAS	**0.032768**
TP53	0.025132
BRAF	**0.023445**
SMAD4	**0.014745**

Some Results

<u>colon</u>	0.061547
folfox	0.032459
woman	0.021636
vemurafenib	0.017319
appreciated	0.014151
cea	0.013780
folfiri	0.013745
braf	0.013290
involving	0.011781
bevacizumab	0.010973

APC	**0.034585**
KRAS	**0.032768**
TP53	0.025132
BRAF	**0.023445**
SMAD4	**0.014745**

Some Results

rectal	0.043272
colorectal	0.031520
adenocarcinoma	0.030553
folfox	0.025718
anal	0.021779
cea	0.013466
rectum	0.013129
colon	0.012224
sooner	0.011782
bevacizumab	0.011074

APC	**0.054803**
KRAS	**0.033947**
TP53	0.019144
ATM	0.008594
MLL2	0.007733

Confounders considered

colon	0.061547
folfox	0.032459
woman	0.021636
vemurafenib	0.017319
appreciated	0.014151
cea	0.013780
folfiri	0.013745
braf	0.013290
involving	0.011781
bevacizumab	0.010973

APC	**0.034585**
KRAS	**0.032768**
TP53	0.025132
BRAF	**0.023445**
SMAD4	**0.014745**

Summary

So far…

- Application of PFM to find associations
- Extension of PFM to deal with covariates
- Semantically meaningful results

Summary

So far...

- Application of PFM to find associations
- Extension of PFM to deal with covariates
- Semantically meaningful results

Future work

- Topic Sparsity
- Non-parametric Extension

Summary

So far…

- Application of PFM to find associations
- Extension of PFM to deal with covariates
- Semantically meaningful results

Future work

- Topic Sparsity
- Non-parametric Extension

For more details and results, please come to my poster!!

Acknowledgements

- Julia Vogt
- Stefan Stark
- Theofanis Karaletsos
- Gunnar Rätsch
- Fernando Perez-Cruz
- Rätsch Lab

European Commission

Machine Learning for Personalized Medicine

1884