

# Aprendizaje Máquina para Medicina Personalizada

Melanie F. Pradier

Universidad Carlos III in Madrid



Madrid, 21 Octubre 2015

# Iniciativa europea Marie-Curie ITN-MLPM

## Aprendizaje Máquina para medicina personalizada



### Machine Learning for Personalized Medicine

Marie-Curie Action: "Initial Training Networks"

- [Home](#)
- [News](#)
- [People](#)
- [Partners](#)
- [Projects](#)
- [Summer Schools](#)
- [Contact](#)
- [Positions](#)
- [Publications](#)

#### About this Network

##### MLPM - Machine Learning for Personalized Medicine

MLPM is a Marie Curie Initial Training Network, funded by the European Union within the 7th Framework Programme. MLMN has started on January 1, 2013 and will be carried out over a period of four years. MLMN is a consortium of several universities, research institutions and companies located in Spain, France, Germany, Belgium, UK, Switzerland, Israel and in the USA. MLMN involves the predoctoral training of 14 young scientists in the research field at the interface of Machine Learning and Medicine. Its goal is to educate interdisciplinary experts who will develop and employ the computational and statistical tools that are necessary to enable personalized medical treatment of patients according to their genetic and molecular properties and who are aware of the scientific, clinical and industrial implications of this research.



From Genetic Data to Medicine



#### Recent Posts

- Prof. Karsten Borgwardt elected Chairman of the 'Machine Learning in Systems Biology' Steering Committee
- 1st MLMN Mini-Hackathon in Basel
- Review, refresh, remember – lectures of Summer School 2014 now online
- New on board: Daniel completes our network
- Milestone reached!

#### Archive

- 2015
  - October (1)
  - April (1)
  - February (2)
- 2014
  - November (2)
  - May (2)
  - April (1)

[www.mlpm.eu](http://www.mlpm.eu)

2/28

# Iniciativa europea Marie-Curie ITN-MLPM

Aprendizaje Máquina para medicina personalizada



**SIEMENS**



 **Inserm**



**ETHzürich**



# Iniciativa europea Marie-Curie ITN-MLPM

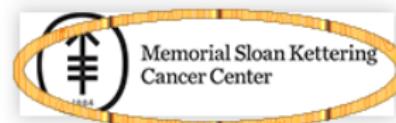
Aprendizaje Máquina para medicina personalizada



**SIEMENS**



 **Inserm**



# Iniciativa europea Marie-Curie ITN-MLPM

## Aprendizaje Máquina para medicina personalizada



**SIEMENS**



# Iniciativa europea Marie-Curie ITN-MLPM

Aprendizaje Máquina para medicina personalizada

## Objetivo

Desarrollo de herramientas estadísticas y computacionales para una medicina personalizada.

## Líneas de investigación

- Descubrimiento de biomarcadores
- Integración de datos
- Mecanismos causales de las enfermedades
- Interacciones entre genes y el entorno

# Iniciativa europea Marie-Curie ITN-MLPM

Aprendizaje Máquina para medicina personalizada

## Objetivo

Desarrollo de herramientas estadísticas y computacionales para una medicina personalizada.

## Líneas de investigación

- Descubrimiento de biomarcadores
- Integración de datos
- Mecanismos causales de las enfermedades
- Interacciones entre genes y el entorno

# Iniciativa europea Marie-Curie ITN-MLPM

Aprendizaje Máquina para medicina personalizada

## Objetivo

Desarrollo de herramientas estadísticas y computacionales para una medicina personalizada.

## Líneas de investigación

- Descubrimiento de biomarcadores
- Integración de datos
- Mecanismos causales de las enfermedades
- Interacciones entre genes y el entorno

# Medicina Personalizada

- ① Diagnósticos precisos
- ② Tratamientos específicos

**Enfoque tradicional "igual para todos"**  
Todos los pacientes con el mismo diagnóstico  
reciben en mismo tratamiento

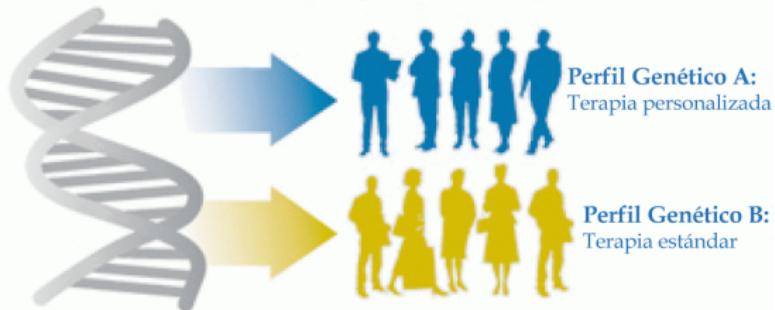


Figure Adaptation from Personalized Medicine Coalition. 2011. What is Personalized Medicine?

# Medicina Personalizada

- ① Diagnósticos precisos
- ② Tratamientos específicos

**Definición:** Forma de la medicina que usa la información de los genes, proteínas y ambiente de una persona para prevenir, diagnosticar y tratar la enfermedad.

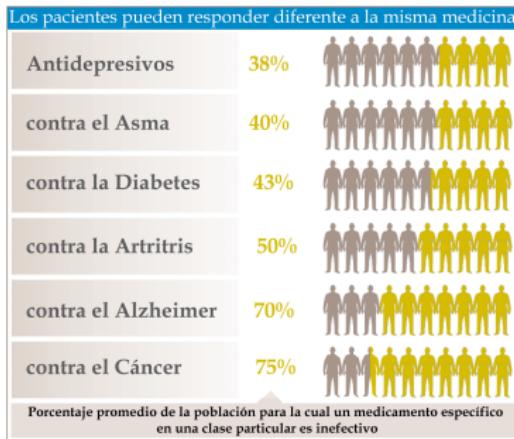


Figure Adaptation from Personalized Medicine Coalition, The Case of Personalized Medicine 2009.

# ¿Por qué ahora?

# ¿Por qué ahora?



# ¿Por qué ahora?

## Fuentes de información

- Secuenciación del genoma
- Sensores de monitorización
- Electronic Health Records

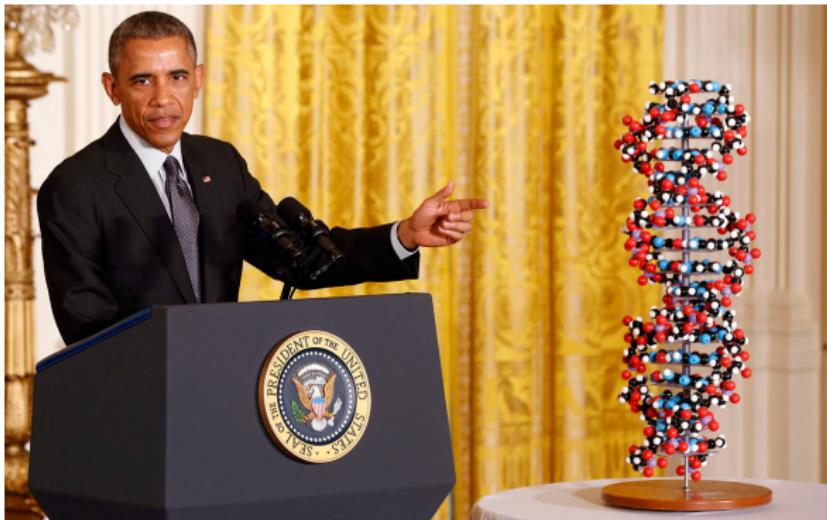


# ¿Por qué ahora?

Datos para EEUU	Hace diez años	Ahora (datos del 2014)
Coste secuenciación genoma	\$22,000,000	\$1000-\$5000
Tiempo secuenciación genoma	2 años	< 1 día
Número de móviles	1M (< 2%)	160M (58%)
Electronic Health Records	20-30%	>90%
Poder de Cómputo	$n$	$n \times 16$

[Francis Collins, The Biology of Genomes, Cold Spring Harbor Laboratory, 5-9 de Mayo 2015]

# ¿Por qué ahora?



- “Precision medicine speech”, 20 de enero de 2015
- Presupuesto de \$215 Millones

# ¿Cómo afrontar tal revolución?

## Solución

- Colaboración multidisciplinar
- Aprendizaje Máquina —> Modelos probabilísticos Bayesianos.



## Desafíos

- Big Data
- ruido
- incompletitud
- complejidad

# ¿Cómo afrontar tal revolución?

## Solución

- Colaboración multidisciplinar
- Aprendizaje Máquina —> Modelos probabilísticos Bayesianos.



## Desafíos

- Big Data
- ruido
- incompletitud
- complejidad

# ¿Cómo afrontar tal revolución?

## Solución

- Colaboración multidisciplinar
- Aprendizaje Máquina —> Modelos probabilísticos Bayesianos.



## Desafíos

- Big Data
- ruido
- incompletitud
- complejidad

# ¿Cómo afrontar tal revolución?

## Solución

- Colaboración multidisciplinar
- Aprendizaje Máquina —> Modelos probabilísticos Bayesianos.



## Desafíos

- Big Data
- ruido
- incompletitud
- complejidad

# ¿Cómo afrontar tal revolución?

## Solución

- Colaboración multidisciplinar
- Aprendizaje Máquina —> Modelos probabilísticos Bayesianos.



## Desafíos

- Big Data
- ruido
- incompletitud
- complejidad

# Ejemplo: diagnóstico médico

- ① Medicina Personalizada
- ② Modelado Bayesiano
- ③ Resultados de Investigación
- ④ Conclusiones

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

## Pregunta

Si la mamografía da positivo, ¿cuál es la probabilidad de que la paciente tenga cáncer?

- 1% ?
- 10% ?
- 70% ?
- 90% ?

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

## Pregunta

Si la mamografía da positivo, ¿cuál es la probabilidad de que la paciente tenga cáncer?

- 1% ?
- 10% ?
- 70% ?
- 90% ?

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

## Pregunta

Si la mamografía da positivo, ¿cuál es la probabilidad de que la paciente tenga cáncer?

- 1% ?
- 10% ?
- 70% ?
- 90% ?

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

$C/\bar{C}$  = tiene o no cáncer

$M/\bar{M}$  = mamografía positiva

o no

- $p(C) = 0.01$
- $p(M|C) = 0.8$
- $p(M|\bar{C}) = 0.096$

$p(C|M)$ ?

Si consideramos 10.000 pacientes....

	$M$	$\bar{M}$
$C$	80	20
$\bar{C}$	950	8950

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

$C/\bar{C}$  = tiene o no cáncer

$M/\bar{M}$  = mamografía positiva

o no

- $p(C) = 0.01$
- $p(M|C) = 0.8$
- $p(M|\bar{C}) = 0.096$

$p(C|M)$ ?

Si consideramos 10.000 pacientes....

	$M$	$\bar{M}$
$C$	80	20
$\bar{C}$	950	8950

# Ejemplo ficticio: diagnóstico médico

## Escenario

- 1% de las mujeres tienen cáncer de mama
- 80% de las mujeres con cáncer de mama dan positivo en el escáner
- 9.6% de las mujeres sanas también dan positivo

$C/\bar{C}$  = tiene o no cáncer

$M/\bar{M}$  = mamografía positiva

o no

- $p(C) = 0.01$
- $p(M|C) = 0.8$
- $p(M|\bar{C}) = 0.096$

$p(C|M)$ ?

Si consideramos 10.000 pacientes....

	$M$	$\bar{M}$
$C$	80	20
$\bar{C}$	950	8950

$$p(C|M) = \frac{p(C,M)}{p(C,M) + p(\bar{C},M)} = \frac{p(C,M)}{p(M)} \simeq 7.8\%$$

# Estadística Bayesiana

- Probabilidad = grado de certidumbre

## Ley de Bayes

$$p(\theta|X) = \frac{p(X|\theta) p(\theta)}{p(X)}$$

- posterior:  $p(\theta|X)$
- likelihood:  $p(X|\theta)$
- prior:  $p(\theta)$
- evidencia:  $p(X)$

# Estadística Bayesiana

- Probabilidad = grado de certidumbre

## Ley de Bayes

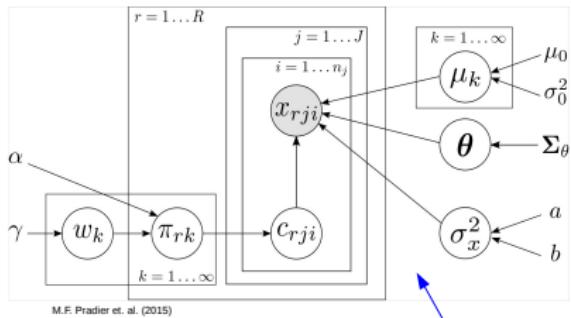
$$p(\theta|X) = \frac{p(X|\theta) p(\theta)}{p(X)}$$

- posterior:  $p(\theta|X)$
- likelihood:  $p(X|\theta)$
- prior:  $p(\theta)$
- evidencia:  $p(X)$

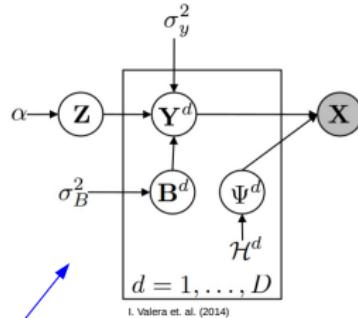
# Modelos Bayesianos

$$p(\theta|X) = \frac{p(X|\theta) p(\theta)}{p(X)}$$

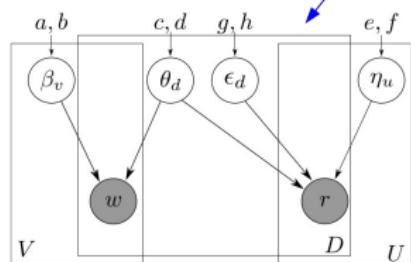
# Modelos Bayesianos



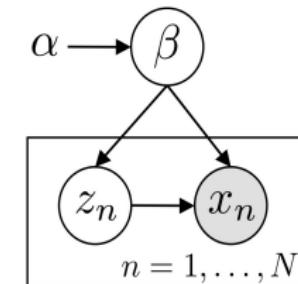
M.F. Pradier et. al. (2015)



$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}$$



P. Gopalan et. al. (2014)

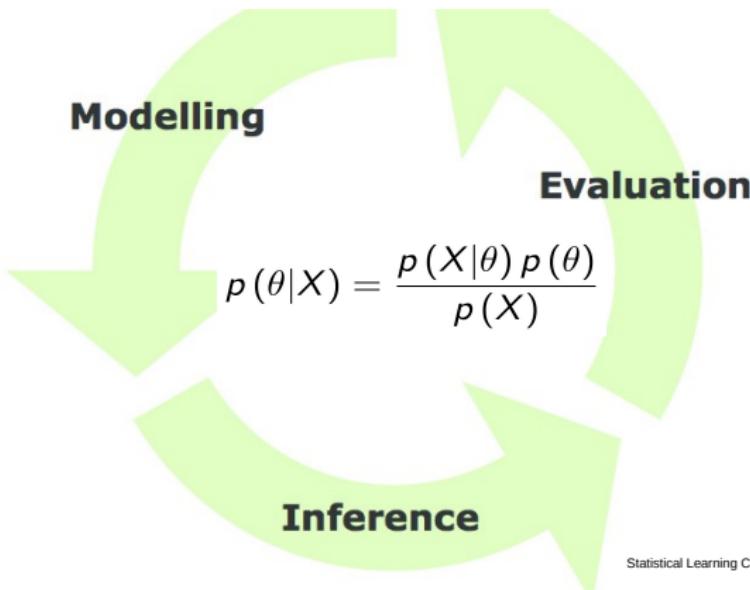


M.F. Pradier et. al. (2014)

# Modelos Bayesianos

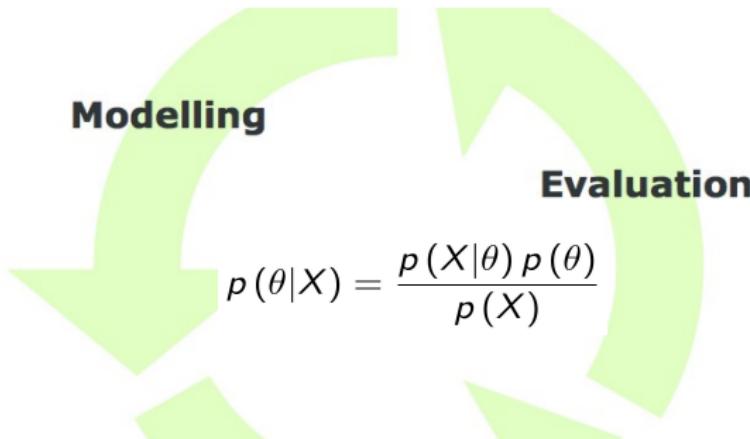
$$p(\theta|X) = \frac{p(X|\theta) p(\theta)}{p(X)}$$

# Modelos Bayesianos



Statistical Learning Cycle [Gelman, 2004]

# Modelos Bayesianos



Factor clave

Colaboración multidisciplinaria entre expertos de distintas áreas.

# Ejemplo: diagnóstico médico

- ① Medicina Personalizada
- ② Modelado Bayesiano
- ③ **Resultados de Investigación**
- ④ Conclusiones

# Asociaciones clínico-genéticas para el cáncer

Colaboración con S. Stark, J.E. Vogt, S. Hyland, T. Karaletsos, y G. Rätsch en el Memorial Sloan-Kettering Cancer Center, New York.

## Objetivo

Descubrir asociaciones genéticas entre fenotipos clínicos y mutaciones genéticas en diversos cancers.

### ¿Por qué importa?

- Mejorar diagnóstico
- Identificar riesgo
- Comprender enfermedad

# Asociaciones clínico-genéticas para el cáncer

Colaboración con S. Stark, J.E. Vogt, S. Hyland, T. Karaletsos, y G. Rätsch en el Memorial Sloan-Kettering Cancer Center, New York.

## Objetivo

Descubrir asociaciones genéticas entre fenotipos clínicos y mutaciones genéticas en diversos cancers.



### ¿Por qué importa?

- Mejorar diagnóstico
- Identificar riesgo
- Comprender enfermedad

# Asociaciones clínico-genéticas para el cáncer

Colaboración con S. Stark, J.E. Vogt, S. Hyland, T. Karaletsos, y G. Rätsch en el Memorial Sloan-Kettering Cancer Center, New York.

## Objetivo

Descubrir asociaciones genéticas entre fenotipos clínicos y mutaciones genéticas en diversos cancers.



## ¿Por qué importa?

- Mejorar diagnóstico
- Identificar riesgo
- Comprender enfermedad

# Asociaciones clínico-genéticas para el cáncer

Colaboración con Memorial Sloan-Kettering Cancer Center, New York.



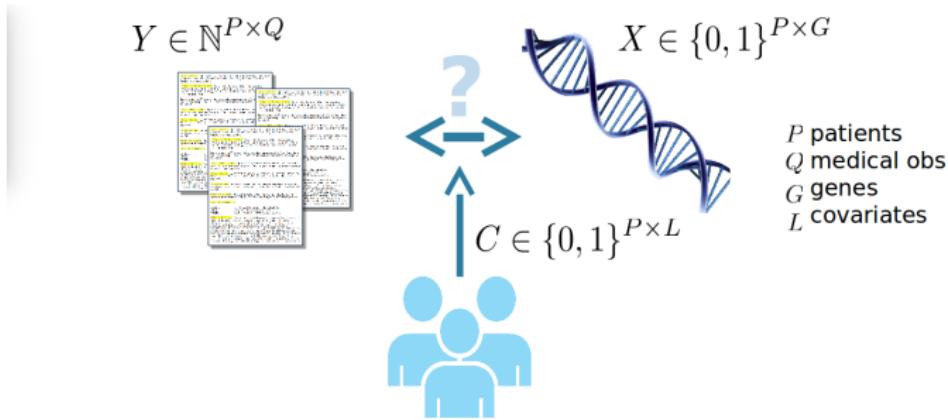
# Asociaciones clínico-genéticas para el cáncer

Colaboración con Memorial Sloan-Kettering Cancer Center, New York.



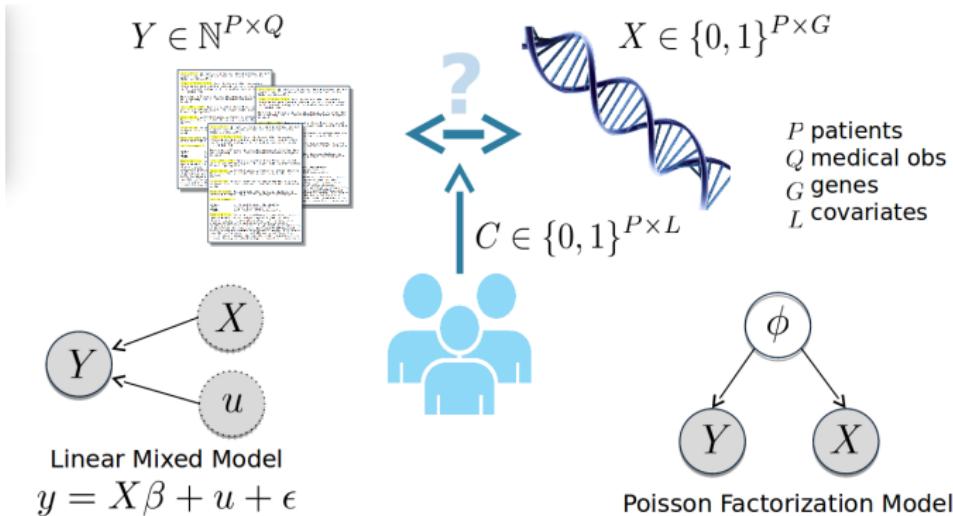
# Asociaciones clínico-genéticas para el cáncer

Colaboración con Memorial Sloan-Kettering Cancer Center, New York.



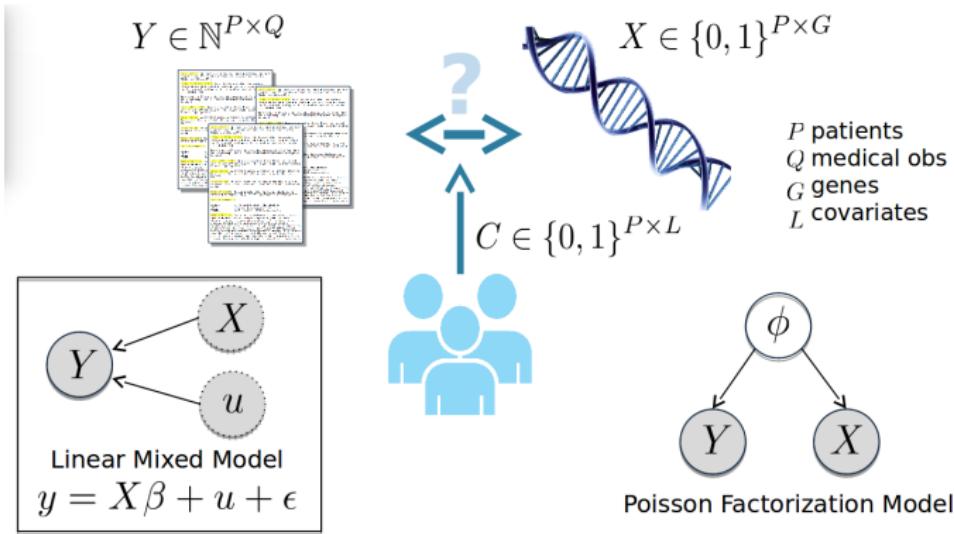
# Asociaciones clínico-genéticas para el cáncer

Colaboración con Memorial Sloan-Kettering Cancer Center, New York.



# Asociaciones clínico-genéticas para el cáncer

Colaboración con Memorial Sloan-Kettering Cancer Center, New York.

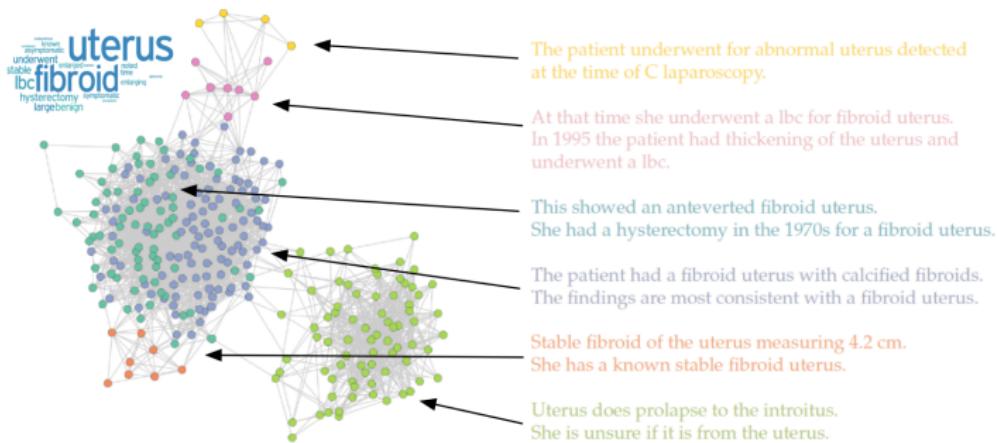


[1] C. Lippert, J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman, “FaST linear mixed models for genome-wide association studies,” *Nat Meth*, vol. 8, no. 10, pp. 833–835, Oct. 2011.

# Asociaciones clínico-genéticas para el cáncer

## Modelos Lineales Mixtos

- Clustering de  $\approx 100M$  de frases
- Algoritmo para análisis de redes sociales (KNN+Louvain hierarchical clustering).



# Asociaciones clínico-genéticas para el cáncer

## Modelos Lineales Mixtos

- Estudio para  $\approx 2000$  pacientes con diversos tipos de cáncer (62 sub-tipos).

Gene	MAF	<i>q</i> -value	$\beta$	Sentence prototype
APC	112	0.0037	0.33	He underwent a colonoscopy which revealed a pedunculated polyp in the ascending colon.
ALK	40	0.0063	0.57	The patient showed a mild decrease in her blood counts.
HNF1A	13	0.0028	0.70	The patient was tearful presented with depressed affect and mood.
TRAF7	11	0.0008	0.59	He has a history of adenoid cystic carcinoma of the salivary gland.
NOTCH2	57	0.0009	0.24	History of multiple nonmelanoma skin cancers and melanoma.
SUFU	14	0.003	0.31	The patient has a solitary fibrous tumor.
ERBB4	93	0.06	0.08	This is a man/lady with metastatic colon carcinoma.

# Asociaciones clínico-genéticas para el cáncer

## Modelos Lineales Mixtos

- Estudio para  $\approx$ 2000 pacientes con diversos tipos de cáncer (62 sub-tipos).

"The APC gene in colorectal cancer"  
(*Eur J Cancer*, 2002 May;38(7):867-71)

" all patients presented with systemic symptoms and signs, including fever, anemia, and thrombocytosis. ... ALK expression was associated with localized disease" (*Mod Pathol* 2003;16(8):828-832)

Genomic Analysis of Non-NF2 Meningiomas Reveals Mutations in **TRAF7**, **KLF4**, **AKT1**, and **SMO** (*Science* 1 March 2013; Vol. 339 no. 6123 pp. 1077-1080)

Gene	MAF	q-value	$\beta$	Sentence prototype
APC	112	0.0037	0.33	He underwent a colonoscopy which revealed a pedunculated polyp in the ascending colon.
ALK	40	0.0063	0.57	The patient showed a mild decrease in her blood counts.
HNF1A	13	0.0028	0.70	The patient was tearful presented with depressed affect and mood.
TRAF7	11	0.0008	0.59	He has a history of adenoid cystic carcinoma of the salivary gland.
NOTCH2	57	0.0009	0.24	History of multiple nonmelanoma skin cancers and melanoma.
SUFU	14	0.003	0.31	The patient has a solitary fibrous tumor.
ERBB4	93	0.06	0.08	This is a man/lady with metastatic colon carcinoma.

Association found in 4 diff. cancers: Bladder Cancer | Head and Neck Carcinoma | Melanoma | Skin Cancer (Non-Melanoma)

ERBB4 is over-expressed in human **colon cancer** and enhances cellular transformation. (*Carcinogenesis* (2015) doi: 10.1093/carcin/bgv049)

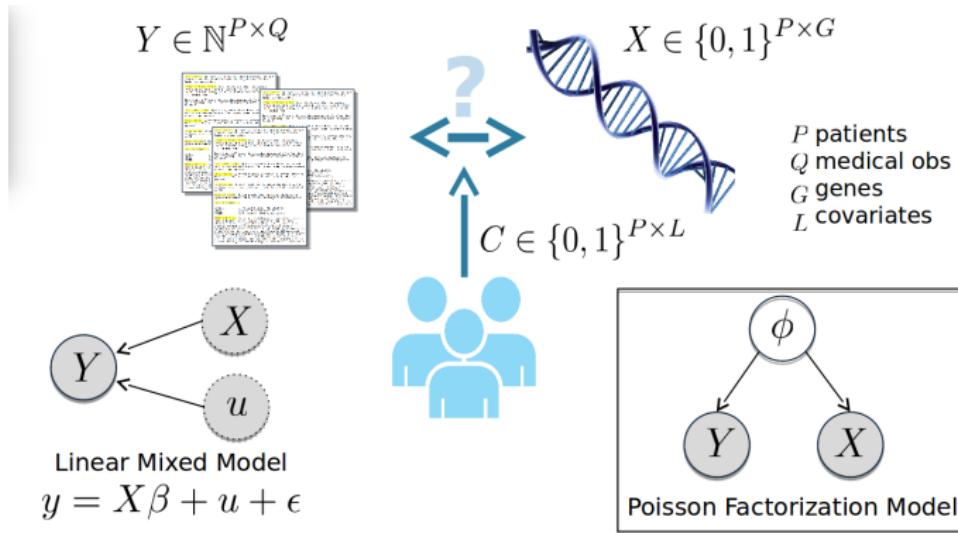
"39 Notch2 promotes **bladder cancer** progression: Pre-clinical rationale for a novel targeted therapy" (*European Urology Supplement* 2014; 1569-9956)

"Evidence for differential expression of Notch receptors and their ligands in melanocytic nevi and cutaneous malignant melanoma" (*Modern Pathology* (2008) 19, 246-254)

M. F. Pradier, S. Stark, S. Hyland, J. E. Vogt, F. Perez-Cruz and G. Rätsch. Large-Scale Sentence Clustering from Electronic Health Records for Genetic Associations in Cancer. Submitted to Workshop on Machine Learning in Computational Biology (MLCB) at NIPS 2015.

# Asociaciones clínico-genéticas para el cáncer

## Poisson Factorization Model

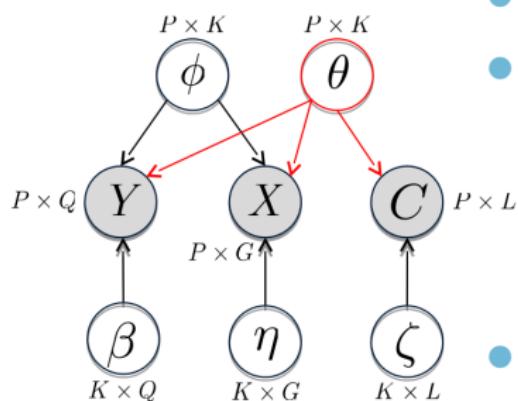


- [1] C. Lippert, J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman, “FaST linear mixed models for genome-wide association studies,” *Nat Meth*, vol. 8, no. 10, pp. 833–835, Oct. 2011.
- [2] P. Gopalan, D. Blei, “Content-based recommendations with Poisson factorization,” presented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

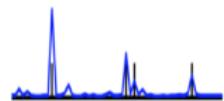
# Asociaciones clínico-genéticas para el cáncer

## Poisson Factorization Model

Poisson Factorization Model



- Clinical factors  $\beta_q$
- Genetic factors  $\eta_g$



$$x_{pq} \sim \text{Poisson}(\phi_p \eta_g + \theta_p \eta'_g)$$

$$y_{pq} \sim \text{Poisson}(\phi_p \beta_q + \theta_p \beta'_q)$$

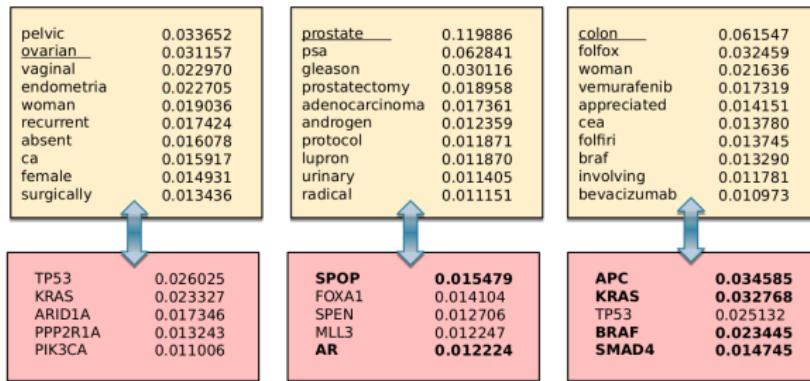
others  $\sim \text{Gamma}(a, b)$

- $\phi$  captures **interesting** structure
- $\theta$  explains away **confounders**

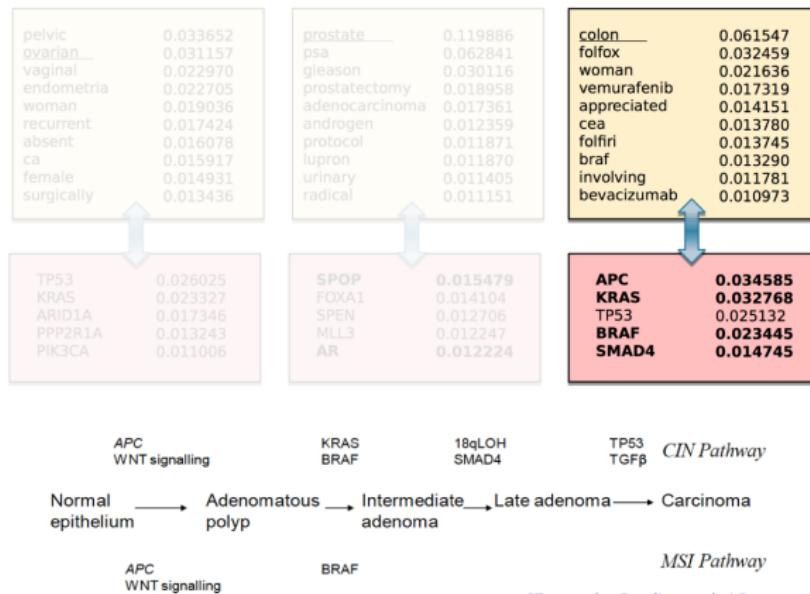
M. F. Pradier, J.E. Vogt, S. Stark, T. Karaletsos, F. Perez-Cruz and G. Rätsch. Probabilistic Analysis of Genetic Associations with Clinical Features in Cancer. Spotlight Talk at the ML Symposium, New York Academy of Sciences, April 2015.

# Asociaciones clínico-genéticas para el cáncer

## Poisson Factorization Model



## Asociaciones clínico-genéticas para el cáncer Poisson Factorization Model



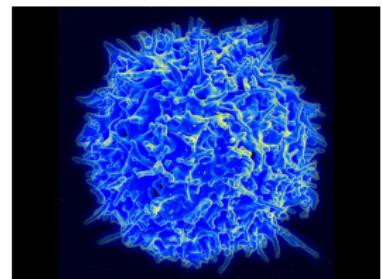
M. F. Pradier, F. Perez-Cruz and G. Rätsch. Bayesian Poisson Factorization Model for Genetic Associations with Clinical Features in Cancer. Working paper. 2015.

# Ensayo clínico de inmunoterapia para cancer de hígado

## Inmunoterapia contra el cáncer

Uso de anticuerpos monoclonales para:

- Destruir células cancerosas
- Frenar su crecimiento
- Radiación dirigida
- Medicamentos locales

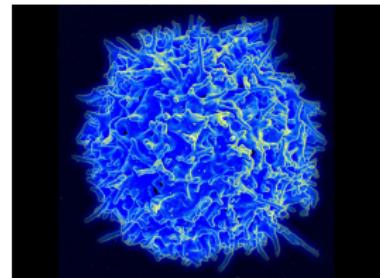


# Ensayo clínico de inmunoterapia para cancer de hígado

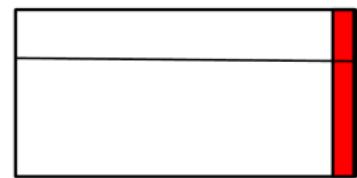
## Inmunoterapia contra el cáncer

Uso de anticuerpos monoclonales para:

- Destruir células cancerosas
- Frenar su crecimiento
- Radiación dirigida
- Medicamentos locales



$N \times D$



$X$

23/28

# Ensayo clínico de inmunoterapia para cancer de hígado

- ① ¿Cómo podemos predecir supervivencia?
- ② ¿Qué características tienen un impacto en supervivencia?  
¿Cómo? → prognostic biomarkers (indep. del tratamiento).
- ③ ¿Qué características ayudan a la acción del medicamento?  
¿Cómo? → predictive biomarkers.

# Ensayo clínico de inmunoterapia para cancer de hígado

- ① ¿Cómo podemos predecir supervivencia?
- ② ¿Qué características tienen un impacto en supervivencia?  
¿Cómo? → prognostic biomarkers (indep. del tratamiento).
- ③ ¿Qué características ayudan a la acción del medicamento?  
¿Cómo? → predictive biomarkers.

# Ensayo clínico de inmunoterapia para cancer de hígado

- ① ¿Cómo podemos predecir supervivencia?
- ② ¿Qué características tienen un impacto en supervivencia?  
¿Cómo? → prognostic biomarkers (indep. del tratamiento).
- ③ ¿Qué características ayudan a la acción del medicamento?  
¿Cómo? → predictive biomarkers.

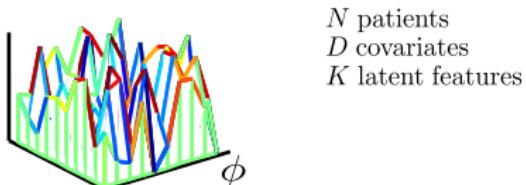
# Biomarker Discovery with the Indian Buffet Process

Collaboration with Francesca Milletti and Oscar Puig at Roche Diagnostics in New York.

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



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



$$Z \sim \text{IBP}(\alpha)$$

- Distribución sobre matrices binarias  $Z_{N \times K}$  donde  $K \rightarrow \infty$

## Fragmentación de la población en subgrupos

- 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)	<b>51P 2.91 1.47</b>	<b>78P 3.36 2.24</b>	<b>23P 3.90 3.84</b>	
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)	<b>09P 4.96 3.96</b>	<b>14P 3.22 1.52</b>	<b>05P 5.59 5.08</b>	
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	

# Conclusiones

En esta charla...

- Medicina Personalizada: es el momento
- Asociaciones clínico-genéticas (colaboración con MSKCC)
- Ensayo clínico de inmunoterapia (colaboración con Roche)

Puntos clave

- ① Colaboración multidisciplinar: factor clave
  - Implicación de los pacientes
  - Comunicación entre expertos
- ② Impacto de la Medicina Personalizada en la sociedad (ética, privacidad, ...)

# Conclusiones

En esta charla...

- Medicina Personalizada: es el momento
- Asociaciones clínico-genéticas (colaboración con MSKCC)
- Ensayo clínico de inmunoterapia (colaboración con Roche)

Puntos clave

- ① Colaboración multidisciplinaria: factor clave
  - Implicación de los pacientes
  - Comunicación entre expertos
- ② Impacto de la Medicina Personalizada en la sociedad (ética, privacidad, ...)

# Reconocimientos

- Fernando Perez-Cruz
- Antonio Artés
- TSC en la UC3M
- Rätsch Lab, New York
- Colaboradores en MSKCC, Roche
- Marie-Curie ITN-MLPM

Gracias!

