



# FÓRMULA DE PREDICCIÓN DE INESTABILIDAD DE MICROSATÉLITES EN EL CÁNCER COLORECTAL EN BASE A PARÁMETROS MORFOLÓGICOS. ESTUDIO MULTICÉNTRICO PROSPECTIVO DE VALIDACIÓN Y OPTIMIZACIÓN DEL MODELO.

R Román<sup>1</sup>, M Verdú<sup>1,3</sup>, A Vidal<sup>3,5</sup>, M Calvo<sup>4</sup>, X Sanjuan<sup>5</sup>, M Jimeno<sup>6</sup>, A Salas<sup>7</sup>, J Autonell<sup>8</sup>, I Trias<sup>9</sup>, M González<sup>1</sup>, B García<sup>1</sup>, N Rodón<sup>1</sup> y X Puig<sup>1,2,3</sup>. <sup>1</sup>BIOPAT. Biopatología Molecular, SL, Grup Assistència; <sup>2</sup>Hospital de Barcelona-SCIAS, Grup Assistència; <sup>3</sup>Histopat Laboratoris; <sup>4</sup>Departament d'Estadística, Universitat de Barcelona; <sup>5</sup>Hospital Universitari de Bellvitge; <sup>6</sup>Hospital del Mar; <sup>7</sup>Hospital Mútua de Terrassa; <sup>8</sup>Hospital General de Vic; <sup>9</sup>Hospital Plató, Barcelona

## INTRODUCCIÓN

Los pacientes con carcinoma de colon o recto, esporádicos o hereditarios, que cursan con alta inestabilidad de microsatélites (MSI-H) presentan mejor pronóstico y peor respuesta a tratamientos basados en 5-fluorouracilo que los originados por la vía clásica de inestabilidad cromosómica (CIN). Por esta razón es importante diferenciar en el momento del diagnóstico ambos tipos de neoplasia. Se han descrito, en los últimos años, algunos caracteres clínico-patológicos que se asocian significativamente a MSI-H, como la localización proximal, la diferenciación mucinosa, el crecimiento en patrón sólido, la presencia de linfocitos intraepiteliales y la reacción inflamatoria "Crohn-like" peritumoral, que pueden ayudar a distinguirlos de los asociados a CIN en el momento del diagnóstico patológico.

En 2005 describimos un modelo logístico basado en variables clínico-patológicas, capaz de predecir la inestabilidad de microsatélites en el carcinoma de colon o recto, con un valor predictivo negativo del 97.8%. El objetivo de éste trabajo ha sido la validación y optimización del modelo.

## MÉTODOS

### Serie de validación:

Nuestra serie de validación consta de 265 tumores primarios de colon o recto, procedentes de cinco laboratorios de Patología de Cataluña, que fueron prospectivamente evaluados siguiendo criterios estándar en los respectivos centros. Del mismo modo, los estudios de inmunohistoquímica (IHQ) fueron realizados siguiendo los protocolos establecidos en cada centro para preservar la variabilidad interobservador. Los bloques de parafina de las muestras tumorales y su correspondiente control de tejido normal se enviaron posteriormente a nuestro laboratorio (Biopat), donde se llevó a cabo el estudio de inestabilidad de microsatélites. Esta serie de validación también incluye 148 casos recopilados y estudiados prospectivamente en nuestro centro.

### Serie de optimización:

La serie global de optimización de nuestro modelo incluye 411 casos de la serie de validación más los 204 casos incluidos en el estudio inicial publicado en 2005. Las variables clinicopatológicas y moleculares evaluadas han sido descritas anteriormente y se encuentran enumeradas en la tabla 1.

### Análisis de inestabilidad de microsatélites:

El estudio MSI se realizó utilizando el panel consenso del NCI compuesto por cinco microsatélites (BAT25, BAT26, D5S346, D2S123 y D17S250) en una PCR multiplex. Conjuntamente también se valoraron seis microsatélites adicionales que habían sido inicialmente diseñados para el estudio de LOH del cromosoma 18q y del 17p. El estado global de inestabilidad fue asignado en base a estos 11 microsatélites.

### Estudio estadístico:

Los análisis estadísticos de validación y homogeneidad entre centros fueron realizados utilizando el test exacto de Fisher para las variables categóricas y mediante el test de Kruskal-Wallis para las variables continuas, implementados respectivamente en los paquetes "stats" y "coin" del entorno estadístico R.

En lugar del proceso clásico de selección conocido como *stepwise regression* con el que se elaboró nuestro modelo original, en el estudio de optimización se ha empleado la metodología descrita por Friedman "*Regularization Paths for Generalized Linear Models via Coordinate Descent*", implementada en el paquete glmnet de R. Para seleccionar las variables que intervienen en el nuevo modelo de regresión logística se ha empleado adicionalmente el criterio de Schwartz y un proceso exhaustivo de remuestreo en el que el elemento básico consiste en generar aleatoriamente dos particiones de la muestra global. En la primera de ellas se seleccionan las variables y se calculan los coeficientes del modelo, mientras que en la segunda partición se comprueba la bondad del ajuste de ese mismo modelo. Se generan estas dos particiones aleatoriamente un total de 1000 veces, anotando qué variables han sido seleccionadas y qué coeficientes de regresión presentan en cada uno de los 1000 diferentes modelos generados. Para confeccionar el modelo final se promedian los coeficientes de las variables incluidas de forma más habitual y se excluyen las menos frecuentes.

## RESULTADOS

### Análisis de validación:

El estudio de homogeneidad entre centros evidenció diferencias significativas en la valoración del patrón de crecimiento, la presencia de respuesta inflamatoria "Crohn-like", el porcentaje de patrón cribiforme y la expresión de Ki-67. A pesar de estas previsibles diferencias, el modelo inicial mantuvo un valor predictivo negativo del 97.8% (Tabla 2).

### Análisis de optimización:

El proceso de remuestreo basado en glmnet demostró que la localización proximal del tumor, un alto porcentaje de patrón sólido y mucinoso extracelular, la presencia de reacción inflamatoria "Crohn-like" y de linfocitos intraepiteliales (TIL) estaban fuertemente asociados a MSI-H, estando estas variables incluidas en un rango entre 976 y 1000 del total de modelos generados. El patrón de crecimiento, presente en 715 modelos, y la expresión de Ki-67 (en 592) y p53 (en 651) también se consideraron asociados al patrón de MSI-H y por consiguiente incluidos en el modelo optimizado que denominamos RERtest8 (Figura 1a). Comparando este modelo con el inicial de 2005, cabe destacar que RERtest8 incluye en la ecuación la presencia de TIL y excluye el porcentaje de patrón cribiforme.

Teniendo en cuenta que la expresión de Ki-67 y p53 son las dos únicas variables que no pueden ser directamente evaluadas en las secciones con hematoxilina-eosina del estudio patológico rutinario y son, además, las que presentan una asociación más débil con el perfil MSI-H, se construyó y evaluó la capacidad predictiva de un segundo modelo optimizado (RERtest6), que no incluye estas dos variables (Figura 1b). Los parámetros estadísticos de ambos modelos optimizados se muestran en la tabla 2. El punto de corte de ambos modelos fue empíricamente establecido en  $p < 0.8$ .

## DISCUSIÓN

La complejidad y el alto coste de los estudios moleculares, junto a la creciente exigencia de información predictiva y la incorporación permanente de nuevas opciones de terapia oncológica personalizada, justifican la búsqueda de modelos de predicción de perfiles moleculares a partir de parámetros clínicos y morfológicos. Este trabajo se ha dedicado a aplicar este objetivo a la inestabilidad de microsatélites en el carcinoma de colon y recto, factor pronóstico y predictivo reconocido de esta neoplasia y asociado, además, al diagnóstico del síndrome de Lynch o cáncer de colon hereditario no asociado a poliposis (HNPCC). El modelo que desarrollamos inicialmente se dirigía, no sólo a identificar presuntos candidatos a HNPCC, sino al diagnóstico específico del fenotipo MSI-H, asociado a mejor pronóstico. El presente estudio ha contribuido a la validación y optimización del modelo.

La fase de validación, con un estudio multicéntrico de 265 tumores, ha demostrado que el modelo inicial es lo suficientemente robusto como para superar las diferencias interobservador y metodológicas entre distintos laboratorios de patología.

El estudio de optimización llevado a cabo con una serie global de 615 casos ha dado como resultado la creación de dos modelos predictivos de MSI-H, RERtest8 y RERtest6, que muestran como diferencia resaltable respecto al modelo inicial la inclusión del TIL en sus ecuaciones a expensas del porcentaje de patrón cribiforme que se excluye. Ambos modelos coinciden con la literatura previa en que la localización proximal del tumor, la presencia de TIL, la reacción inflamatoria "Crohn-like" y el componente mucinoso extracelular están estrechamente ligados al perfil MSI-H. Debe resaltarse, no obstante, que el impacto de cada uno de estos parámetros es distinto en cada modelo. La localización proximal y la presencia de reacción inflamatoria "Crohn-like" son las variables predictivas más destacables en nuestros modelos, seguidas por la presencia de TIL y del patrón de crecimiento expansivo. Este último no se encuentra incluido en ninguno de los modelos anteriormente publicados por otros autores. Otro factor a destacar es que la edad del paciente no resulta una variable predictiva ni en nuestro modelo inicial ni en los modelos optimizados RERtest8 y RERtest6. Esto puede ser debido a una baja proporción de casos hereditarios en nuestra serie, que presenta una mediana de edad de 69,2 años.

La prioridad general de nuestro estudio fue la maximización del valor predictivo negativo para conseguir una herramienta predictiva capaz de reducir el número de casos que requerirían análisis molecular de MSI y, al mismo tiempo, de fácil implementación a la rutina diaria. El resultado son los dos modelos, RERtest8 y RERtest6, el último sin parámetros de IHQ, con valores predictivos negativos del 97.93% y 97,91%, respectivamente que permiten reducir el número de casos que requieren estudio molecular al 10% de los carcinomas de colon o recto estudiados en el laboratorio de Patología. Asumiendo como adecuado el valor predictivo negativo alcanzado, sólo los casos clasificados como MSI-H requerirían confirmación.

Este nuevo recurso, en manos del patólogo general, incorpora en la sistemática de diagnóstico la predicción del perfil molecular del carcinoma de colon (MSI-H vs CIN) a la vez que, junto a los protocolos clínicos disponibles, contribuye a la identificación de pacientes con HNPCC.

## REFERENCIAS

- Colomer A. et al. A novel logistic model based on clinico-pathological features predicts microsatellite instability in colorectal carcinomas. *Diagn Mol Pathol* 2005;14:213–23.
- Jenkins MA. et al. Pathology features in Bethesda Guidelines predict colorectal cancer microsatellite instability: a population-based study. *Gastroenterology* 2007;133:48-56.
- J. Friedman, T. Hastie and Rob Tibshirani, Glimnet: Lasso and elastic-net regularized generalized linear models, (2008), <http://www-stat.stanford.edu/~hastie/Papers/glmnet.pdf>
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- G. Schwarz, Estimating the Dimension of a Model, *Annals of Statistics*. Volume 6, Number 2 (1978), 461-464.

| Variables Categóricas                 | MSS (%)<br>n=563 | MSI-H (%) n=52 |
|---------------------------------------|------------------|----------------|
| <b>Genero</b>                         |                  |                |
| Hombres                               | 336 (59.7)       | 22 (42.3)      |
| Mujeres                               | 227 (40.3)       | 30 (57.7)      |
| <b>Localización</b>                   |                  |                |
| Proximal                              | 177 (31.4)       | 45 (86.5)      |
| Distal                                | 386 (68.6)       | 7 (13.5)       |
| <b>Configuración</b>                  |                  |                |
| Exofítico                             | 244 (43.3)       | 23 (44.2)      |
| Ulcerado                              | 230 (40.9)       | 23 (44.2)      |
| Estenosante                           | 89 (15.8)        | 6 (11.5)       |
| <b>Grado de invasión</b>              |                  |                |
| pT1                                   | 37 (6.6)         | 1 (1.9)        |
| pT2                                   | 83 (14.7)        | 7 (13.5)       |
| pT3                                   | 271 (48.1)       | 33 (63.5)      |
| pT4                                   | 172 (30.6)       | 11 (21.2)      |
| <b>Invasión vascular intramural</b>   |                  |                |
| Presente                              | 172 (30.6)       | 16 (30.7)      |
| Ausente                               | 391 (69.4)       | 36 (69.2)      |
| <b>Invasión vascular extramural</b>   |                  |                |
| Presente                              | 141 (25.0)       | 11 (21.2)      |
| Ausente                               | 422 (75.0)       | 41 (78.9)      |
| <b>Invasión venosa intramural</b>     |                  |                |
| Presente                              | 40 (7.1)         | 2 (3.8)        |
| Ausente                               | 522 (92.7)       | 50 (96.2)      |
| <b>Invasión venosa extramural</b>     |                  |                |
| Presente                              | 113 (20.1)       | 8 (15.4)       |
| Ausente                               | 450 (79.9)       | 44 (84.6)      |
| <b>Invasión perineural intramural</b> |                  |                |
| Presente                              | 54 (9.6)         | 2 (3.8)        |
| Ausente                               | 509 (90.4)       | 50 (96.2)      |
| <b>Invasión perineural extramural</b> |                  |                |
| Presente                              | 76 (13.5)        | 3 (5.8)        |
| Ausente                               | 487 (86.5)       | 49 (94.2)      |
| <b>Patrón de crecimiento</b>          |                  |                |
| Expansivo                             | 193 (34.3)       | 32 (61.5)      |
| Infiltrativo                          | 370 (65.7)       | 20 (38.5)      |
| <b>Reacción "Crohn like"</b>          |                  |                |
| Presente                              | 190 (33.7)       | 41 (78.8)      |
| Ausente                               | 373 (66.3)       | 11 (21.2)      |
| <b>TIL</b>                            |                  |                |
| Presente                              | 86 (15.3)        | 29 (55.8)      |
| Ausente                               | 477 (84.7)       | 23 (44.2)      |
| <b>Adenoma residual</b>               |                  |                |
| Presente                              | 176 (31.3)       | 15 (28.8)      |
| Ausente                               | 387 (68.7)       | 37 (71.2)      |

| Variables numéricas            | Media ± DS  | Media ± DS  |
|--------------------------------|-------------|-------------|
| Edad                           | 69.1 ± 11.6 | 69.8 ± 12.9 |
| Tamaño tumoral (mm)            | 43.2 ± 21.1 | 57.5 ± 17.1 |
| Patrón sólido (%)              | 5.0 ± 12.9  | 23.7 ± 35.2 |
| Patrón mucinoso (%)            | 9.2 ± 20.7  | 32.3 ± 32.5 |
| Patrón cribiforme (%)          | 4.3 ± 10.1  | 12.3 ± 19.7 |
| Patrón micropapilar (%)        | 2.0 ± 6.9   | 0.3 ± 1.5   |
| Patrón microglandular (%)      | 3.2 ± 9.8   | 0.6 ± 2.2   |
| Afectación linfática (n)       | 2.0 ± 3.7   | 2.2 ± 5.7   |
| Índice proliferativo Ki-67 (%) | 62.4 ± 21.7 | 72.6 ± 17.0 |
| Sobreexpresión p53 (%)         | 43.3 ± 38.5 | 17.9 ± 24.1 |

Tabla 1: Características clinicopatológicas y moleculares de la serie de optimización. MSS: casos estables, MSI-H: casos con alta inestabilidad de microsatélites.

|                           | Validación | RERtest8 | RERtest6 | Estudio Inicial |
|---------------------------|------------|----------|----------|-----------------|
| Exactitud                 | 91.32      | 92.48    | 92.09    | 95.10           |
| Sensibilidad              | 66.67      | 78.14    | 78.01    | 77.78           |
| Especificidad             | 93.44      | 93.80    | 93.39    | 96.77           |
| Valor predictivo positivo | 46.67      | 53.45    | 51.78    | 70.00           |
| Valor predictivo negativo | 97.02      | 97.93    | 97.91    | 97.83           |
| Total (N)                 | 265        | 615      | 615      | 204             |

Tabla 2: Parámetros estadísticos de la serie de validación y los modelos optimizados.

**A: RERTest 8**

$$1 - \frac{1}{1 + e^x} < 0.8 \text{ MSI-H}$$

a= Localización: 0= proximal, 1=distal  
b= Crecimiento: 0= expansivo, 1=infiltrativo  
c= Crohn: 0=presente, 1=ausente  
d= TIL %  
e= Patrón sólido %  
f= Patrón mucinoso %  
g= Ki-67%  
h= sobreexpresión p53 %

$$x = (2.925 \cdot (a^*1.563) + (b^*1.034) + (c^*1.426) + (d^*0.588) - (e^*0.046) - (f^*0.028) - (g^*0.031) + (h^*0.021))$$

**B: RERTest 6**

$$1 - \frac{1}{1 + e^x} < 0.8 \text{ MSI-H}$$

a= Localización: 0= proximal, 1=distal  
b= Crecimiento: 0= expansivo, 1=infiltrativo  
c= Crohn: 0=presente, 1=ausente  
d= TIL %  
e= Patrón sólido %  
f= Patrón mucinoso %

$$x = (1.114 \cdot (a^*1.523) + (b^*1.046) + (c^*1.591) + (d^*0.893) - (e^*0.042) - (f^*0.028) - (g^*0.031))$$

Figura 1: Modelos optimizados: A: RERtest8 y B: RERtest6