



Combinación de datos clínico-epidemiológicos y datos genéticos del microorganismo en pacientes sometidos a ventilación mecánica: utilidad de un método exploratorio de datos

A.Lacoma^{1,4}, M.Gomes-Fernandes^{1,4,5}, E.Mesalles², F.Arméstar², R.Villar^{1,4}, I.Casas³, S.Molinos^{1,4}, M.Giménez^{1,4}, N.Pagán¹, C.Prat^{1,4}.

1. Servei de Microbiologia. Hospital Universitari Germans Trias i Pujol. Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol. Universitat Autònoma de Barcelona. 2. Servicio de Medicina Intensiva. Hospital Universitari Germans Trias i Pujol. Universitat Autònoma de Barcelona. 3. Servicio de Medicina Preventiva. Hospital Universitari Germans Trias i Pujol. Universitat Autònoma de Barcelona. 4. CIBER Enfermedades Respiratorias. 5. CAPES Foundation, Ministry of Education of Brazil, Brasília.

Objetivos: 1). Evaluar un análisis exploratorio de datos en una cohorte de pacientes sometidos a ventilación mecánica con aislamiento de *Staphylococcus aureus* en muestra respiratoria y 2). Realizar un análisis exhaustivo que permita identificar asociaciones relevantes entre todas las variables recogidas, así como en función de los diferentes grupos de estudio.

Material y Métodos: Previo acuerdo de confidencialidad se combinaron en un único fichero, los **datos clínico-epidemiológicos** provenientes de una base de datos SPSS con los **datos genéticos del microorganismo** que estaban en formato excel. Estos últimos se obtuvieron a partir de un *microarray* de DNA que detecta la presencia/ausencia de distintos genes codificantes para factores de virulencia. Los datos se combinaron, analizaron y exploraron automáticamente mediante el software "AutoDiscovery" (Butler Scientifics S.L., www.butlerscientifics.com), con el objetivo de conocer qué factores del paciente y qué genes del microorganismo caracterizan mejor a la cohorte de pacientes en función del grupo de estudio, la persistencia en el aislamiento pese a recibir tratamiento ajustado a antibiograma (>72h) y el desarrollo de complicaciones respiratorias, incluyendo la mortalidad. Los grupos de estudio fueron: pacientes con colonización bronquial, traqueobronquitis, y neumonía. También se incluyó un grupo de portadores nasales ingresados en UCI en el mismo periodo como grupo control. El análisis estadístico se hizo con el software SPSS v15.

Resultados:

- El EDA permite explorar diferentes y múltiples comparaciones entre las distintas variables de estudio, ordenándolas según la relevancia y su presencia/ausencia en los diferentes grupos de estudio (Figura 1).
- Algunas de las asociaciones resultantes confirmaron correlaciones previamente obtenidas en el análisis estadístico univariante y multivariante realizado con métodos estadísticos contrastados. En la Figura 2 se muestra un ejemplo de resultado confirmado con el método EDA, que muestra que la persistencia en el aislamiento pese a ser más frecuente en aquellos casos con aislamiento de *S.aureus* metilicilin resistente (MRSA), también se observa en cepas sensibles (MSSA).
- Por otro lado también permite generar y extraer relaciones no hipotetizadas en un primer momento seleccionando las variables de interés (Figura 3). En caso de que exista alguna relación, está debe confirmarse posteriormente.

Evaluated at	Indep. Var.	Depend. Var.	From	To	Strength	# Samples	Filter 1 - Variable	Filter 1 - Value
17/11/2015 16:40:18	[chronic_respiratory_disease]	days_mv_until_first_isolation	No	Yes	25 %	123		
17/11/2015 16:52:00	[presence_of_comorbidities]	glasgow	No	Yes	25 %	27	study_group	Pneumonia
17/11/2015 16:48:59	[sex]	days_persistence_recod	Male	Female	25 %	22	study_group	Colonizat
17/11/2015 16:39:51	[blood_culture]	icu_length_stay	Not done	Positive	25 %	134		
17/11/2015 16:39:14	[microorganism_recod]	icu_length_stay	Altres	Altres BGN	25 %	43		
17/11/2015 16:49:50	[ab_after_culture_result]	total_days_mv	No	Yes	25 %	46	study_group	Colonizat
17/11/2015 16:39:40	[connective_sample]	icu_length_stay	No	Yes	25 %	134		
17/11/2015 16:39:04	[inmed_isolation_first_isolate]	crp	Yes	No	25 %	146		
17/11/2015 16:40:18	[chronic_respiratory_disease]	only_icu_daysmv	No	Yes	24 %	148		
17/11/2015 16:52:04	[ita]	days_mv_until_first_isolation	No	Yes	24 %	27	study_group	Pneumonia
17/11/2015 16:49:47	[previous_ab]	days_mv_until_first_isolation	No	Yes	24 %	48	study_group	Colonizat
17/11/2015 16:51:46	[sex]	total_days_mv	Female	Male	24 %	27	study_group	Pneumonia
17/11/2015 16:51:46	[sex]	icu_length_stay	Female	Male	24 %	27	study_group	Pneumonia
17/11/2015 16:49:00	[study_group]	icu_length_stay	Carriage	Pneumonia	24 %	134		
17/11/2015 16:49:00	[inmed_isolation_first_isolate]	days_persistence_recod	No	Yes	24 %	22	study_group	Colonizat
17/11/2015 16:42:54	[cmv_virus]	leucocytes	No	Yes	24 %	50	study_group	Colonizat
17/11/2015 16:46:00	[presence_of_comorbidities]	glasgow	No	Yes	24 %	43	study_group	Tracheobro
17/11/2015 16:49:47	[previous_ab]	total_days_mv	No	Yes	24 %	48	study_group	Colonizat
17/11/2015 16:51:47	[inmed_isolation_first_isolate]	leucocytes	Yes	No	24 %	20	study_group	Pneumonia
17/11/2015 16:51:46	[sex]	days_mv_until_first_isolation	Female	Male	24 %	27	study_group	Pneumonia
17/11/2015 16:41:22	[other_cis_involvement]	only_icu_daysmv	No	Yes	24 %	148		

Figura 1. Imagen con resultados después de realizar EDA.

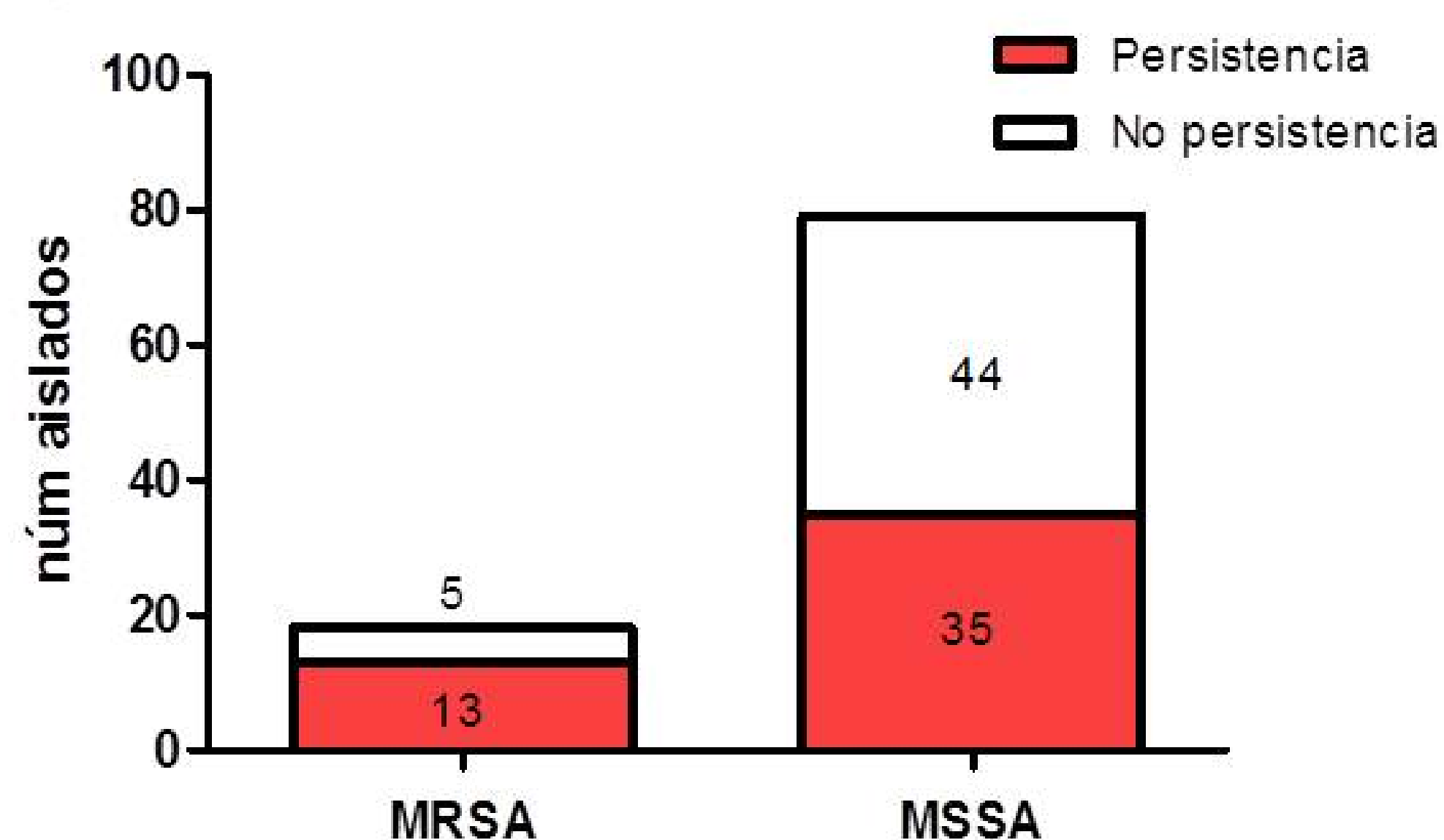


Figura 2. Distribución de las cepas (MRSA y MSSA) en función de si se aislaron en casos persistentes o no persistentes (p=0.039).

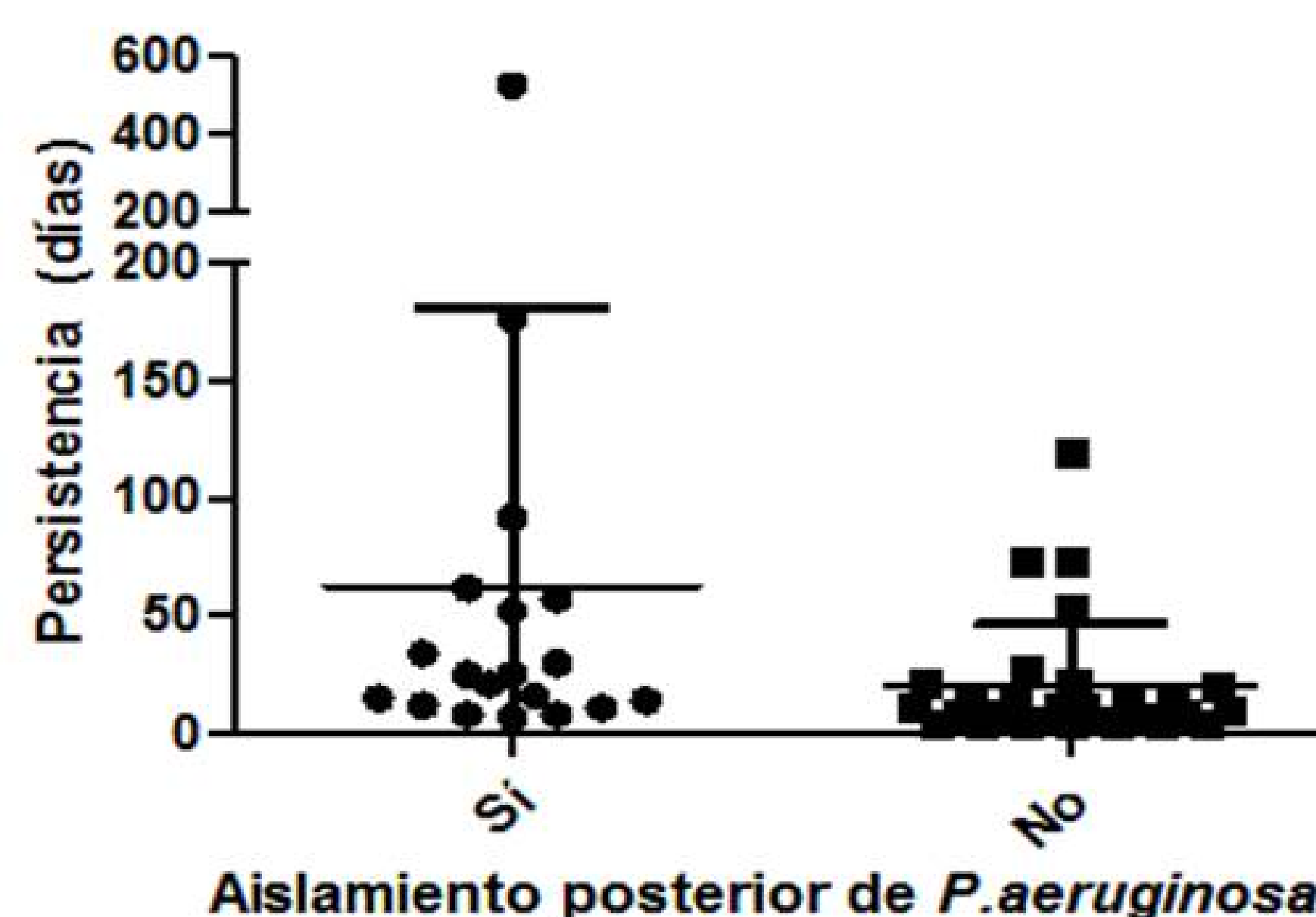


Figura 3. Días de persistencia en función de si hubo aislamiento posterior de *P.aeruginosa* al considerar los tres grupos de estudio (p= 0.0086).

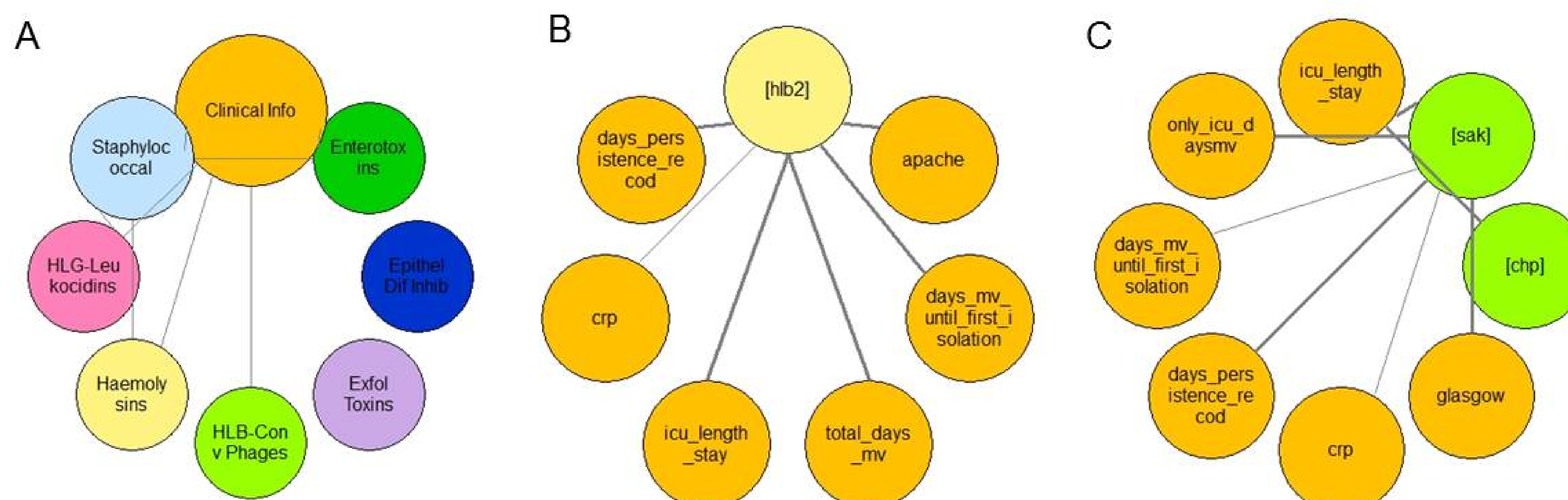


Figura 4. A. Representación de las relaciones existentes entre las sondas incluidas en el array de DNA y el conjunto de variables clínicas. B. Relaciones existentes entre el gen *hlyB* (perteneciente al grupo de hemolisinas) y las variables clínicas. C. Relaciones existentes entre los genes *sak* y *chp* (*hlyB* converting phages) y las variables clínicas.

- Se analizaron en detalle las relaciones más relevantes entre el gen *hlyB* y las variables clínicas (Figura 4). Se detectaron más días de ventilación mecánica, mayor estancia en UCI y más días transcurridos desde el inicio de la intubación hasta el primer aislamiento en el grupo de pacientes con neumonía. En todos los casos, el aumento de días se correspondió con la presencia del gen.
- Por otro lado, se analizaron las relaciones existentes con los genes pertenecientes al grupo *hlyB* converting bacteriophages, destacando el gen *chp* y *sak*. La presencia del gen *chp* se asoció a una mayor estancia en UCI, aunque solamente en el grupo de pacientes portadores. En cambio, la presencia del gen *sak* se relacionó a una mayor estancia en UCI en el grupo de pacientes con traqueobronquitis.

Conclusiones: El software AutoDiscovery es una herramienta útil para extraer, de forma eficiente, relaciones entre bases de datos complejas con un número de variables amplio que puedan ser confirmadas posteriormente y que mejoran las conclusiones e impacto del trabajo. Las herramientas de análisis exploratorio de datos facilitan la identificación de patrones y correlaciones inesperadas.

Financiación: SEPAR 054/2011, FIS PI13/0148. IP:C.Prat