

Genómica comparativa de genes implicados en la Enfermedad de Parkinson

Ignacio Marín

Instituto de Biomedicina de Valencia

CSIC

Genómica comparativa

- Deriva de un campo clásico: evolución molecular
- Todo aquello que no se puede conocer mediante el análisis de genes o genomas aislados
 - *Sensu stricto*: Comparación de genomas
 - Extracción de información detallada sobre familias génicas, genes, productos génicos... *basada en la comparación de genomas* → Contexto en esta charla

Para qué sirve: algunas ideas

- **Definición de organismos modelo** que posean genes ortólogos al que nos interesa
- **Predicción de funciones** de genes o productos génicos, mediante comparación con datos de otras especies
- **Predicción del efecto de mutaciones:** mediante comparación de secuencias en diversas especies, modelado tridimensional basado en datos de otras especies, etc.

Estrategias de análisis

- Se precisan procedimientos capaces de:
 - **manipular y ordenar** grandes cantidades de información
 - **proporcionar una visión integrada** de la información

Algunos ejemplos...

	80	*	100	*	120	*	140	
AB062092h	: DAEQPNNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
NM_103396	: DAEQPNNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 79
AC018929c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073163a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 97
AC006918b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AB073156d	: DEQELPNNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046433c	: DEQELPNNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 45
AF296829a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 40
AL161509c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AF262042d	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046433e	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AF296826b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AB046435b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AB073166e	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AB062091	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073155d	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC068901	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC079028a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AF296828a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
NM_122817	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046439a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046426b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC069329a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AF296827a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046429a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
NM_122815	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046428c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046428b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC011621b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046428e	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC011621e	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046428d	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC063973a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 101
AB046436a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
AF147259c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
NM_116846	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
AL161507a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC069554a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046425a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC069552a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC066259a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
NM_126815	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073197a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC007918b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073158c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 101
AB073157a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 101
AC018660c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
AC019012b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
AL391731b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 101
AL138643a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 101
AC083859f	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB062092g	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046430b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 100
AF296827b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 100
AC0639731	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: -
AB046436h	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 67
NM_103359	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC020646a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB062087b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB062087a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB046426a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB024037a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AF104920a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AL161504d	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AF147261b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073158b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC069555a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073157b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AB073166d	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 102
AC004483b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 82
AC006259c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 99
AF272709b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AB005249	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
NM_126917	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AC006446a	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AC006446b	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AF077407	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AC006446c	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66
AC005965	: DEQEQPTNTGASDFHNNHQRHG		IVPPVQNNNFEEKSLIAMVQGNKHF		GLMEDPLDHLDEF			: 66

Comparación de secuencias

Relativamente estandarizada

Pueden analizarse miles de secuencias

Conservación/cambio aminoácido

Similitud (cualitativa) entre secuencias

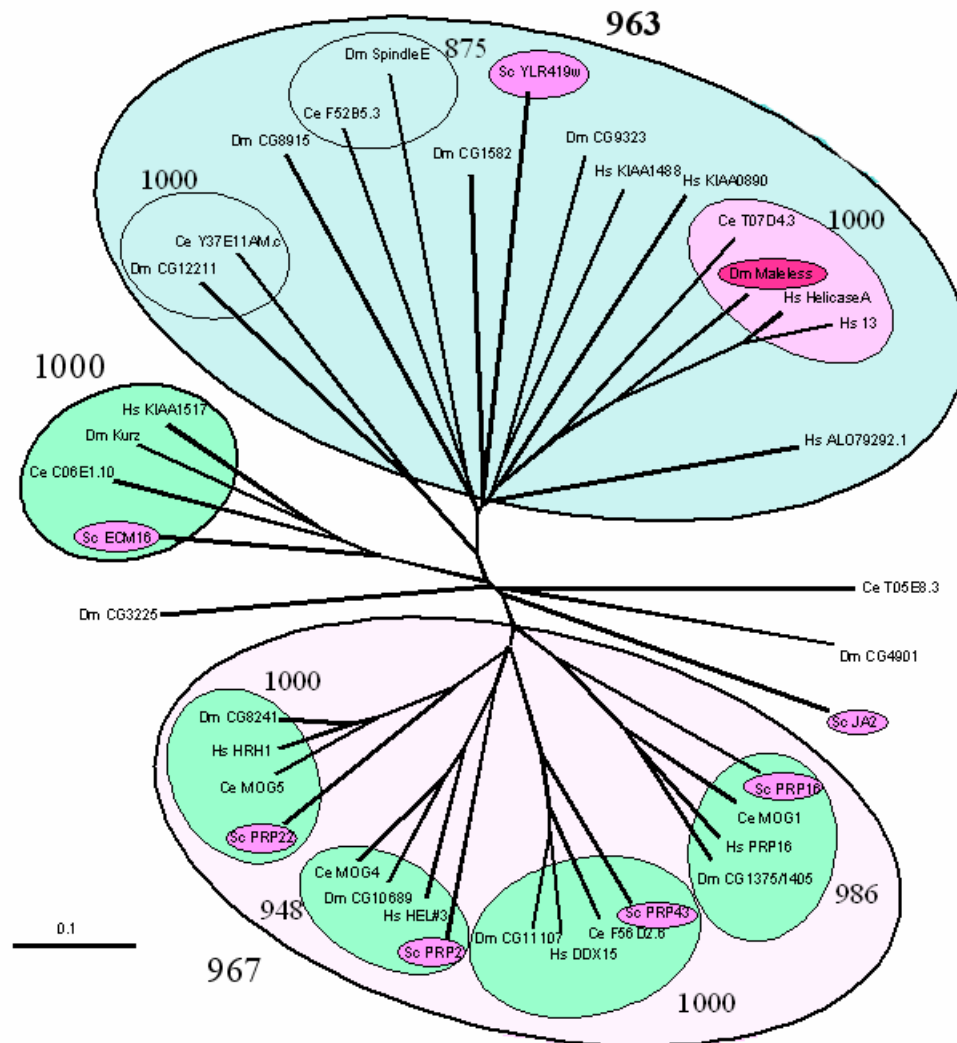
Análisis filogenético

Relativamente estandarizado
(si $N < 500$)

Diversos métodos

Similitud entre secuencias

Relaciones ortología/paralogía



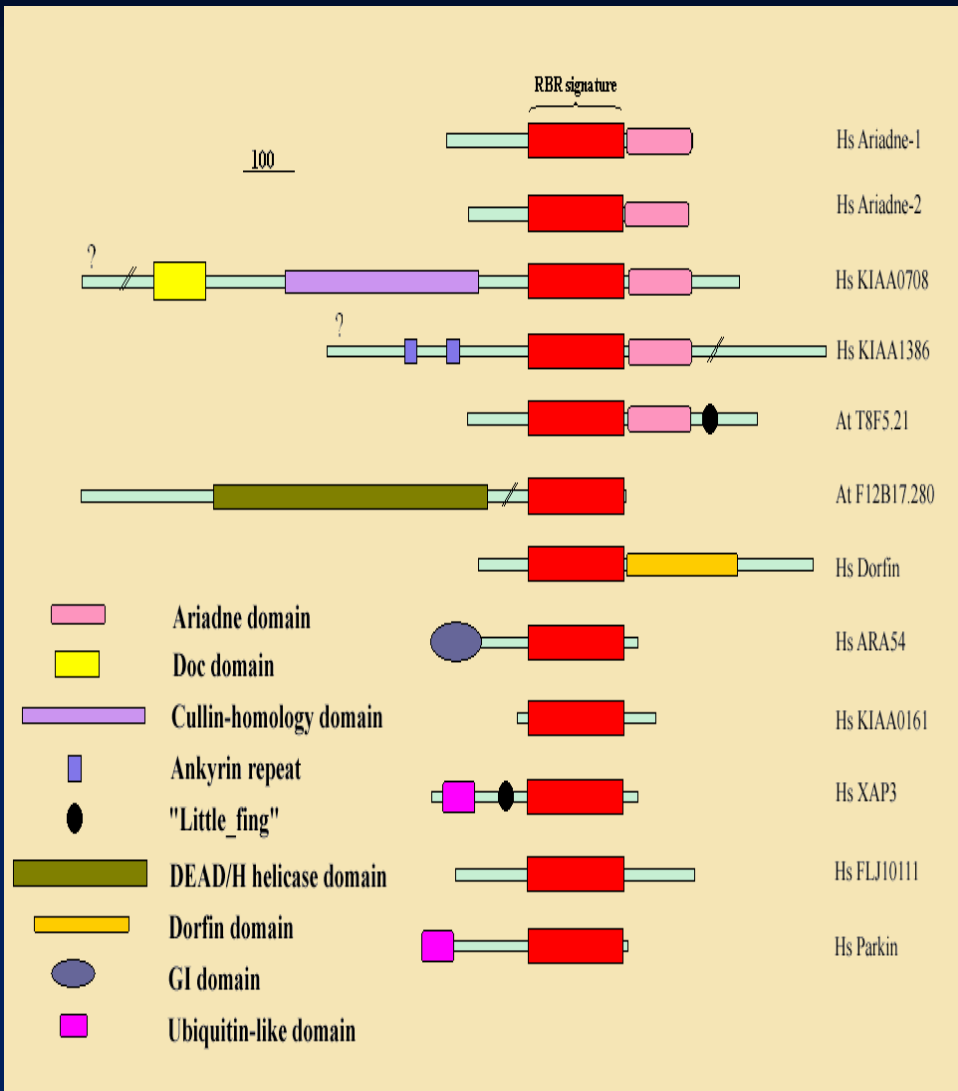
Análisis estructural: dominios proteicos

En general, sencillo

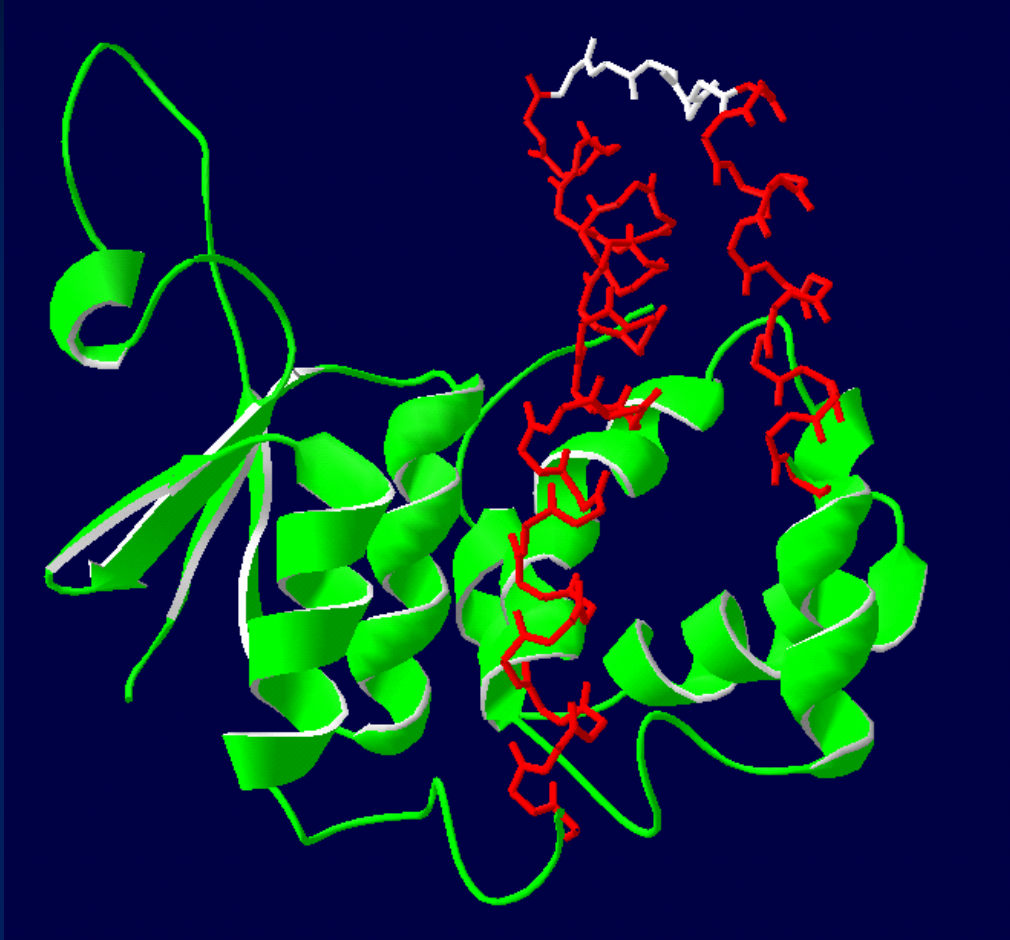
Favorecido por la existencia de
excelentes bases de datos
específicas

Similitud estructural

Deducción de funciones



Análisis estructural: modelización de estructuras 3D (entre especies)



Puede ser posible o no

Bases de datos específicas

*Conservación de funciones,
efecto de mutaciones, etc.*

Análisis comparativo de patrones de interacción proteína-proteína

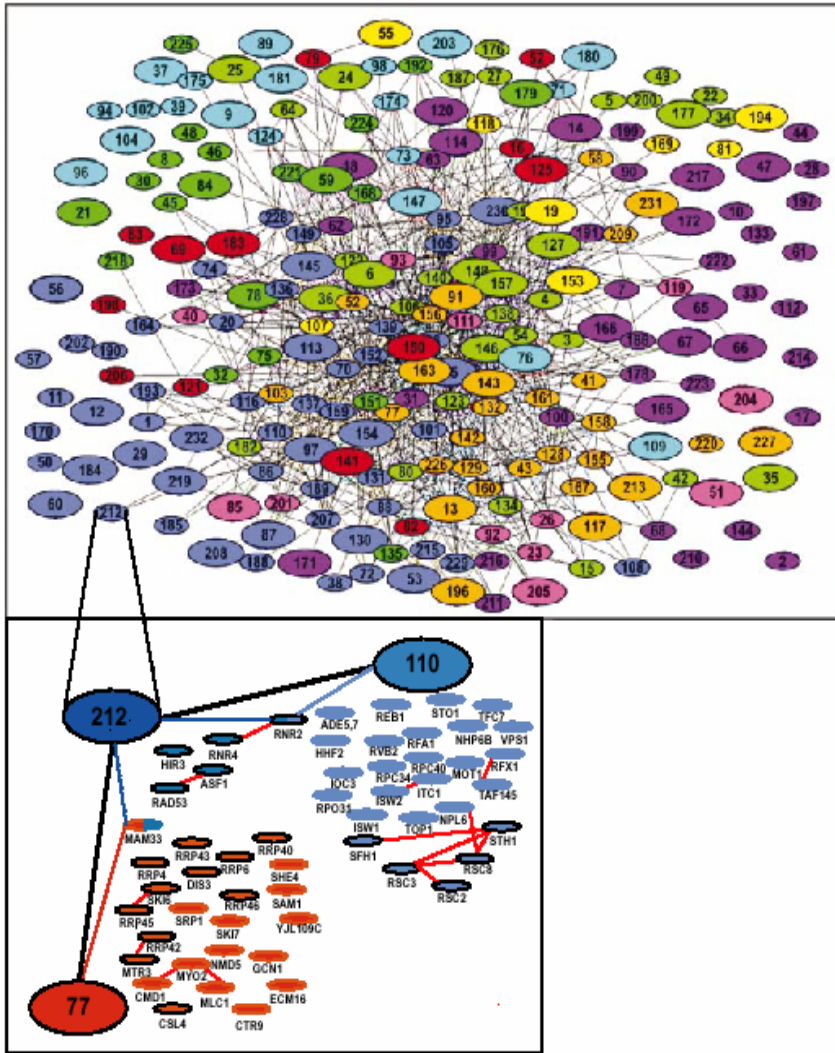
Métodos en desarrollo

Bases de datos específicas

Restringido a algunas especies

Fiabilidad de datos obtenidos a partir de análisis masivos: 50 – 70%

Función de proteínas, estructura funcional de la célula completa

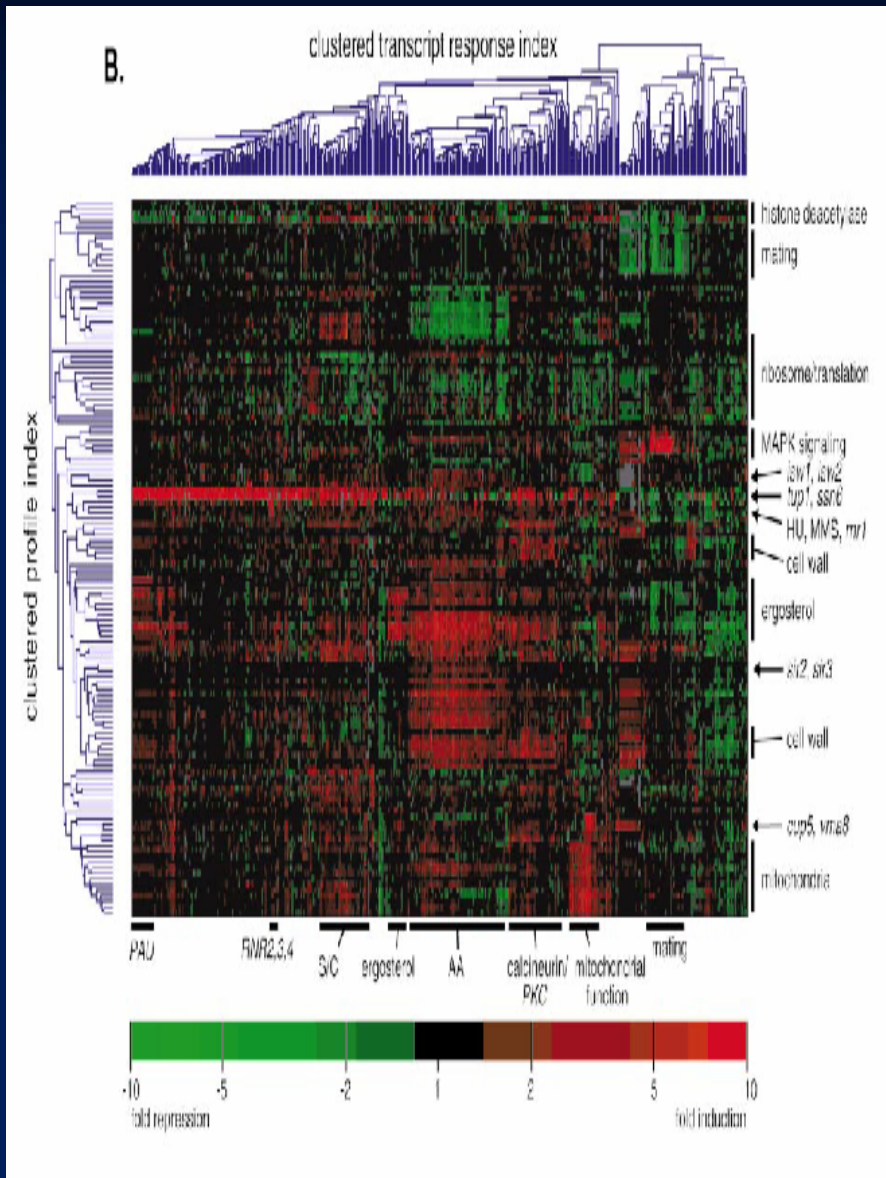


Análisis comparativo de patrones de expresión

En proceso de estandarización

Complicaciones serias en el tratamiento de los datos: ruido, problemas estadísticos asociados a muchos datos/hipótesis, etc.

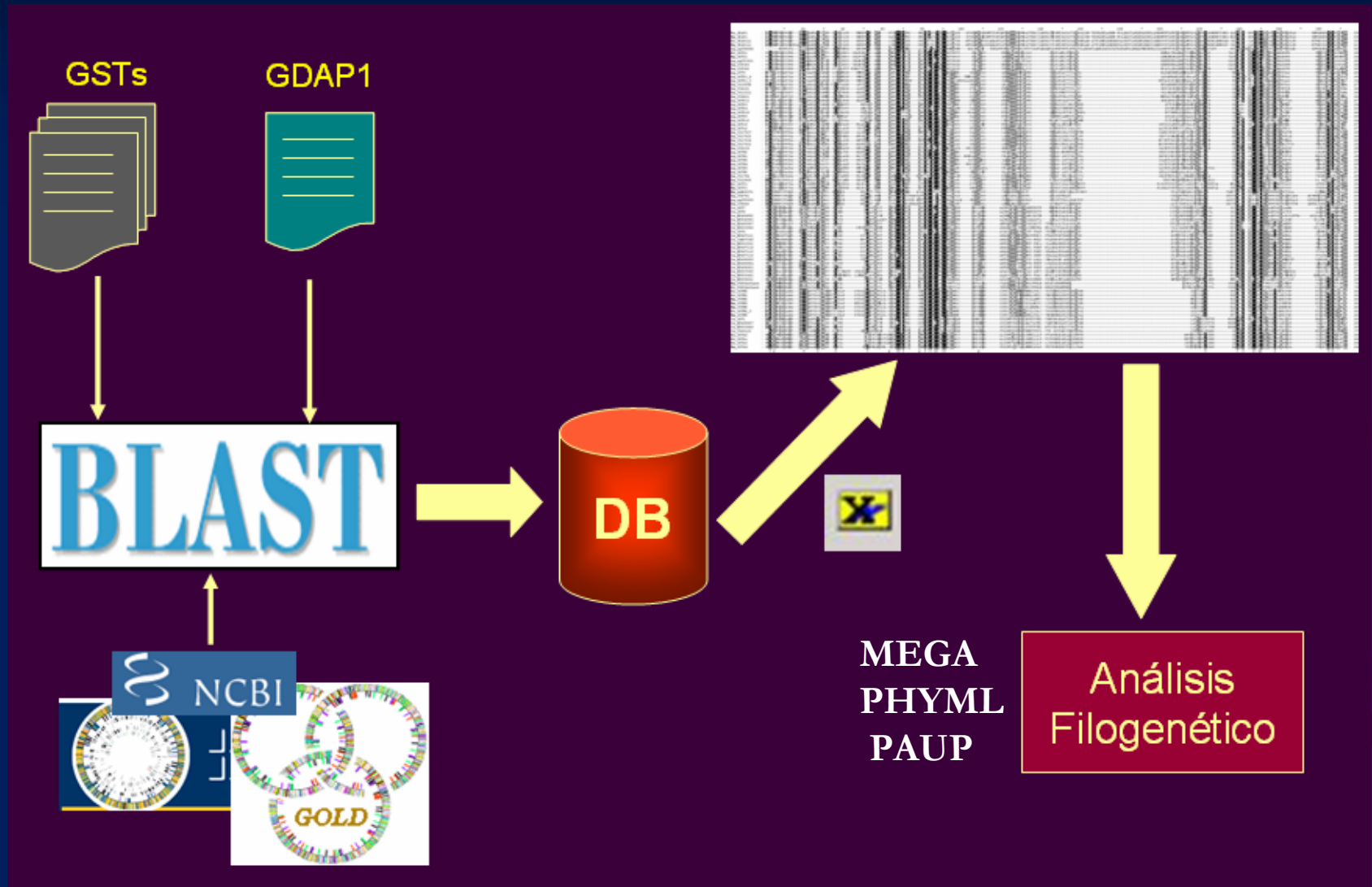
Asignación de funciones mediante correlación de la expresión, módulos funcionales,...



“Pipelines”

- Se deben organizar las diversas herramientas para llevar a cabo análisis estandarizados con un fin definido
- “Pipeline Función”: caracterización de funciones en un contexto evolutivo → **INTEGRACIÓN**
 - Selección de secuencias: BLASTP, TBLASTN,...
 - Alineamiento de secuencias: CLUSTALX, Muscle, T-Coffee...
 - Visualización alineamientos: GeneDoc
 - Análisis filogenéticos: MEGA4, PHYML, PAUP...
 - Estructuras (dominios, secundarias, 3D): InterProScan, Swiss-Model, SwissPDBViewer...
 - Interactoma (análisis, visualización): UVCLUSTER, TreeTracker, Pajek
 - Expresión (análisis, visualización): Cluster, TreeView...

Ejemplo



Ejemplo

Análisis Estructural de Proteínas

```
>Analysis AP00062
REEVARQKRVIAEEQHHEAPFEAVGEEVEIIO
DTFTPVQRAHHPAAKREPILBEYQYLKELKFQTRY
P
HKETMRLELQGDMEYLMELALATIMSCQCDQY
K*EGCQLLATLKVHFEVDBBEKEIGQGVITIKVRB
VEYTLDIRLSDIFQFPGQGIQGYD*EELKBUWNTI
ADPKPISPTKSKSBSIHHPVIRHLRQCAITLFPKKAT
RFVQEEELCMIQDSLVFILRETKDQRRLAQDRASTBL
SNVLLSHLLSYKEAATIHRSQIHDSLCVQDILLTPILO
VVMMEFQTPDVLPKFIDLBYLKQKDFLEKTLADRYL
FKFHHBELOPSLLPQCEHRTLVKTRAHIMPFPBAL
L
HIDQDAREDEPEYSPAQYEQHEVITQADYEEQIEFE
QHIDQPEQQTSEDEQAHYEEQAQDEQEAQVHKK
LQVLEELQMLYHKTMDQFKKKMRTMKK*LEBMAA
QLKDFQSKPRPPBSHSHHSGSTGAARRAERIDQ
PR
AGBFE
```

Estructuras Secundarias

SSPRO
SOPM
Prof 1.0

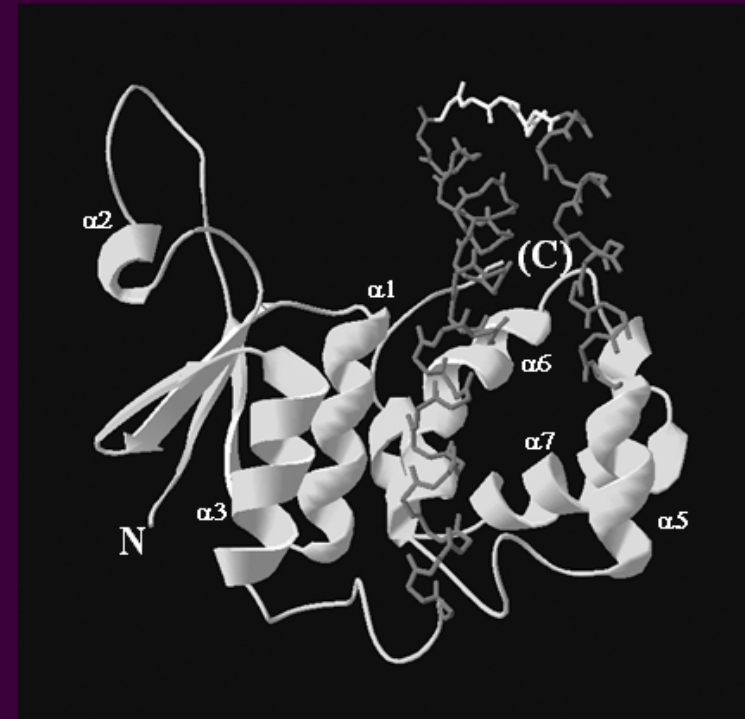
Motivos Funcionales

TMpred
TMHMM
PairCoil

Estructura 3D

Swiss-Prot
Modeller

Visualización



Ejemplo: genes implicados en la enfermedad de Parkinson

- Segunda más frecuente enfermedad neurodegenerativa
- Significativa influencia genética en la probabilidad de sufrir Parkinson esporádico
- Casos infrecuentes de Parkinson familiar asociados a mutaciones en genes concretos

Genes implicados en la enfermedad de Parkinson

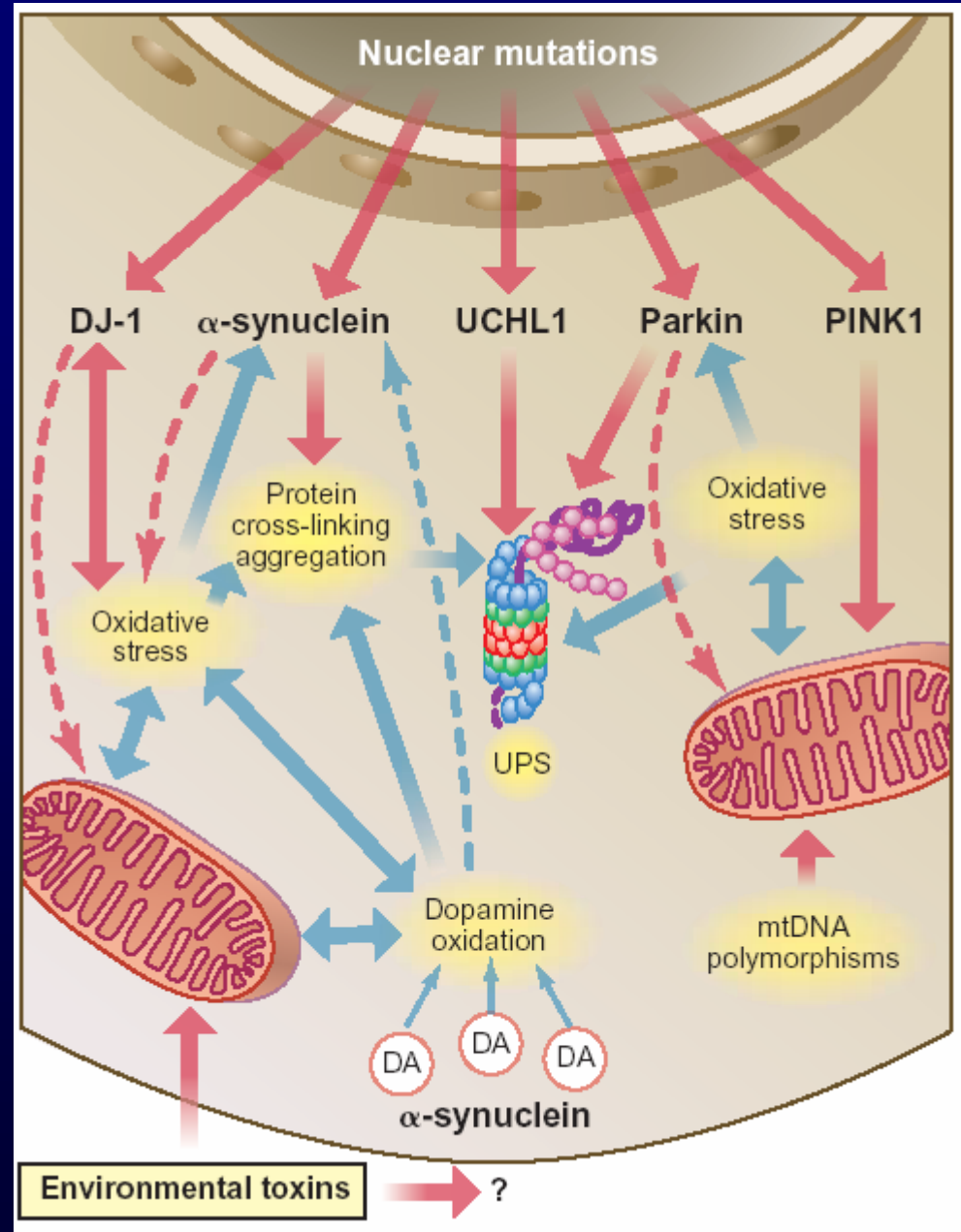
Gen	Cromosoma	Herencia	Características fenotípicas	C. de Lewy
<i>SNCA</i> (PARK1)	4q21	Aut. Domin.	Temprano, demencia	SI
<i>Parkin</i> (PARK2)	6q25	Aut. Reces.	Parkinson juvenil	NO (si)
PARK3	2p13	Aut. Domin.	Tardío, Demencia	SI
<i>UCHL1</i> (PARK5)	4p14	Aut. Domin. (?)	Tardío	?
<i>PINK1</i> (PARK6)	1p35-36	Aut. Reces.	Temprano	?
<i>DJ-1</i> (PARK7)	1p36	Aut. Reces.	Temprano	?
<i>LRRK2</i> (PARK8)	12p11.2-q13.1	Aut. Domin.	[esporádico]	NO/SI
<i>ATP13A2</i> ((PARK9)	1p36	Aut. Reces.	Kufor-Rakeb	?
PARK10	1p32	Aut. Reces. (?)	Tardío	?
PARK 11	2q36-37	Aut. Domin.	-	?
<i>HTRA2</i> (PARK13)	2p12	Influye en susceptibilidad Parkinson esporádico		
<i>NR4A2/NURR1</i>	2q22-23	Aut. Domin.	Tardío	?
<i>MAPT</i>	17q21	Influye en susceptibilidad Parkinson esporádico		

Los papeles de las proteínas codificadas por los genes implicados en Parkinson son muy poco claros

¿Regulación del metabolismo proteico?

¿Respuesta al estrés oxidativo?

Los datos comparativos pueden ayudar a comprender sus funciones



TÍPICO EJEMPLO DE INTEGRACIÓN

Los hallazgos de partida (ca. 1999):

- *Parkin*: nuevo gen implicado en Parkinson familiar
- La proteína Parkin humana posee unas regiones características, ricas en Cys, His
- ¿FUNCIÓN?

Muchas proteínas relativamente parecidas a Parkin en diversos organismos

gi 31248006 ref XP_316606.1 	ENSANGP00000011460 [Anopheles ...	91	6e-17
gi 29125038 emb CAD52894.1 	ARIADNE-like protein ARI12 [Ara...	91	7e-17
gi 38090302 ref XP_146805.2 	similar to XY body protein [Mu...	91	7e-17
gi 47212510 emb CAF93732.1 	unnamed protein product [Tetrao...	90	1e-16
gi 17550524 ref NP_509382.1 	zn-finger, RING and zn-finger,...	90	1e-16
gi 17550522 ref NP_509383.1 	zn-finger, RING and zn-finger,...	90	1e-16
gi 39583782 emb CAE74855.1 	Hypothetical protein CBG22707 [...	90	1e-16
gi 29570509 gb AAC91744.1 	Hypothetical protein C17H11.6c [...	90	1e-16
gi 25406076 pir F06193	hypothetical protein [imported] - A...	89	2e-16
gi 12847362 dbj BAE27541.1 	unnamed protein product [Mus mu...	89	2e-16
gi 9910592 ref NP_064396.1 	ring finger protein 14; TRIAD2 ...	89	2e-16
gi 26334073 dbj BAC30754.1 	unnamed protein product [Mus mu...	89	3e-16
gi 38565956 gb AAH62374.1 	IBRDC3 protein [Homo sapiens]	88	5e-16
gi 47224268 emb CAG09114.1 	unnamed protein product [Tetrao...	88	6e-16
gi 41053706 ref NP_957431.1 	similar to IBR domain containi...	87	6e-16
gi 48145521 emb CAG32983.1 	RNF14 [Homo sapiens]	87	7e-16
gi 35493782 ref NP_056250.3 	ring finger protein 19; double...	87	7e-16
gi 14916696 sp Q9NV58 RN19_HUMAN	RING finger protein 19 (Do...	87	7e-16
gi 4757762 ref NP_004281.1 	ring finger protein 14 isoform ...	87	8e-16
gi 7305437 ref NP_038951.1 	ring finger protein (C3HC4 type...	87	9e-16
gi 31981877 ref NP_666154.2 	IBR domain containing 2; cDNA ...	87	1e-15
gi 47230084 emb CAG10498.1 	unnamed protein product [Tetrao...	87	1e-15
gi 47077394 dbj BAB18585.1 	unnamed protein product [Homo s...	86	2e-15
gi 34906822 ref NP_914758.1 	similar to pre-mRNA splicing f...	86	2e-15
gi 34577096 ref NP_899645.1 	ring finger protein 14 isoform...	86	3e-15
gi 7023254 dbj BAA1900.1 	unnamed protein product [Homo sa...	84	6e-15
gi 38089927 ref XP_146907.3 	similar to XY body protein [Mu...	84	7e-15
gi 38017613 ref NP_542130.1 	ring finger protein 144; ubiqu...	83	1e-14
gi 47217895 emb CAG05017.1 	unnamed protein product [Tetrao...	83	1e-14
gi 7710078 ref NP_057903.1 	parkin [Mus musculus] >gi 54569...	83	1e-14
gi 40788887 dbj BAA11478.2 	KIAA0161 [Homo sapiens]	83	2e-14
gi 11464987 ref NP_064478.1 	Parkin [Rattus norvegicus] >gi...	82	4e-14
gi 15238647 ref NP_200833.1 	zinc finger (C3HC4-type RING f...	81	5e-14
gi 15242882 ref NP_201178.1 	IBR domain-containing protein ...	81	5e-14
gi 1723116 sp P50876 U714_HUMAN	Ubiquitin conjugating enzym...	81	7e-14
gi 38045938 ref NP_055561.2 	ring finger protein 144; UbcM4...	80	7e-14
gi 20385796 gb AAH21456.1 	parkin isoform [Rattus norvegicus]	80	1e-13
gi 7717035 gb AAF68666.1 	parkin [Rattus norvegicus] >gi 11...	80	1e-13
gi 20385788 gb AAH21452.1 	parkin isoform [Rattus norvegicus]	80	1e-13
gi 17537627 ref NP_496609.1 	zn-finger, cysteine-rich C6HC...	80	1e-13

BLAST con Parkin

Similitud significativa
pero escasa

Parece circunscribirse
a varias regiones
concretas

¿Cómo se interpreta?

Se detecta un dominio llamado IBR



The screenshot shows the NCBI BLAST interface. At the top, the NCBI logo is on the left, and "formatting BLAST" is on the right. Below the logo, there are tabs for "Nucleotide", "Protein", "Translations", and "Retrieve results for an RSD". The "Protein" tab is selected. The main text says "Your request has been successfully submitted and put into the Blast Queue." Below that, it says "Query = (503 letters)". A horizontal bar represents the protein sequence, with a red box labeled "IBR" indicating a detected domain. A red arrow points from the top right towards the "IBR" label. Below the bar, it says "Putative conserved domains have been detected, click on the image below for detailed results." At the bottom, there is a text box for "The request ID is" with the value "1087919935-19820-45974696973.BLASTQ4". Below this, there are buttons for "Format" and "Result".

NCBI *formatting* **BLAST**

Nucleotide Protein Translations Retrieve results for an RSD

Your request has been successfully submitted and put into the Blast Queue.

Query = (503 letters)

Putative conserved domains have been detected, click on the image below for detailed results.

IBR

The request ID is 1087919935-19820-45974696973.BLASTQ4

Format or Result

The results are estimated to be ready in 51 seconds but may be done sooner.

¡Una proteína con un único dominio!
(totalmente falso, tiene 4)

Más análisis automatizados de estructuras: resultados ambiguos

SMART

References: Sönnichsen et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 2837-2844
Lemstra et al. (2004) *Nucleic Acids Res* 32, D142-D144

HOME | SETUP | LITERATURE | ABOUT | GLOSSARY | WHAT'S NEW | FEEDBACK

Simple
Modular
Architecture
Research
Tool

Your sequence is identical to [swissprot|Q94981|ARI1_DROME](#), displaying precalculated results.

Due to overlapping domains, there are 3 representations of the protein

1 100 200

RING – RING ???

RING – IBR (¡Todas estas estructuras resultan ser falsas!)

RING- IBR- IBR

Sin análisis comparativos no estamos yendo a ninguna parte

Análisis comparativos: primera parte

- Tomar todas las secuencias relacionadas con Parkin obtenidas con análisis tipo BLAST
- Alinear con ClustalX
- Visualizar regularidades con un editor de secuencias (GeneDoc)
- Análisis filogenéticos (ClustalX, Mega)

Tienen extensos patrones de His, Cys conservadas, muy característicos

	*	20	*	40	*	6
SpAL360054	:	CEICYDEGC-----LPFFSAEC-----DHEFOLACYRQ-YLDSRISEGES--	:	39		
ScYKR017c	:	CIICCDKKD-----TETFALEC-----GHEYQINCYRH-YIKDKLHEGNI--	:	39		
CandidaCon	:	CFICCESYPK-----TTVYSLTC-----GHQFCFSCYQQ-YIGNEIVRGEL--	:	40		
Hs/MmARI1	:	CQICYLNYPN-----SYFTGLEC-----GHKFCMCWSE-YLTTKIMEEGM--	:	40		
DmARI1	:	CEICFSQLPP-----DSMAGLEC-----GHRFCMPQWHE-YLSTKIIVAEGL--	:	40		
DmCG12362	:	CGICFCSC-----DELIGLGC-----GHNFOAACWKQ-YLANKTCSEGL--	:	38		
CeC27A12.6	:	CDICCSM-----DELSGLSC-----NHRACAEQWQA-YLTNKIVSDAQ--	:	37		
CeC27A12.7	:	CDICCSM-----DELSGLSC-----NHRACAEQWQA-YLTNKIVSDAQ--	:	37		
CeC27A12.8	:	CDICCSL-----GELSGLSC-----NHRACQEWKA-YLTNKIANNNAQ--	:	37		
CeY73F8A.3	:	CDVCCSM-----TRLSGLAC-----AHRACDEQWKA-YLTEKIVDVGQ--	:	37		
CbriggsaeA	:	CQICME-----GRLTGPAQ-----NHKASEQWKA-YVTEKIK-EGQ--	:	36		
HsARI2	:	CAVOMQFVRK-----ENLLSLAC-----QHOFORSCWEQ-HCSVLVKDGVG--	:	40		
MmARI2	:	CAVOMQFVRK-----ENLLSLAC-----QHOFORSCWEQ-HCSVLVKDGVG--	:	40		
DmARI2	:	CPVCASSQLG-----DKFYSLAC-----GHSFKDCWTI-YFETQIFQGIS--	:	40		
CeT12E12.1	:	CSVCAMDGY-----TELPHLTC-----GHCFOEHCWKS-HVESRLSEGVA--	:	39		
AtMGF10.12	:	CDVOMEDDLPS-----NVMTRMEC-----GHRFONDQWIG-HFTVKINEGES--	:	41		
AtMGF10.13	:	CDICMEEDLSK-----YAMTRMEC-----GHRFONDQWKE-HFTVRIINEGEG--	:	41		
AtF7H1.11	:	CDIOVEDVPG-----YQLTRMDC-----GHSFONNCWTG-HFTVKINEGQS--	:	40		
AtF10M10.1	:	CDVOMEDLPG-----DHMTRMDC-----GHCFOFNCWTE-HFTVQINEGQS--	:	40		
AtF24D7.x	:	CGICFESYPL-----EETISVSC-----GHPFOATCWTG-YISTSIINDGPG--	:	40		
AtT8F5.21	:	CGICFETFLS-----DKLHAAAC-----GHPFQDSWEG-YITTAINDGPG--	:	40		
AtF20M17.2	:	CGICFESYTR-----KEIASVSC-----GHPYOKTCWTG-YITTKIEDGPG--	:	40		
AtF20M17.1	:	CGICFESYTR-----KEIARVSC-----GHPYOKTCWTG-YITTKIEDGPG--	:	40		
AtF20M17.1	:	CGICFESYTR-----EEIARVSC-----GHPYOKTCWAG-YITTKIEDGPG--	:	40		
AtT20M3.15	:	CGACGESHPH-----KNLASVSC-----GHRIOTRCWTS-HINKIISEKPAA--	:	41		
HsKIAA1386	:	CDICMCSISVF-----EDPVDMPQ-----GHDFORGCWES-FLNLKIQEGEA--	:	41		

C C

C

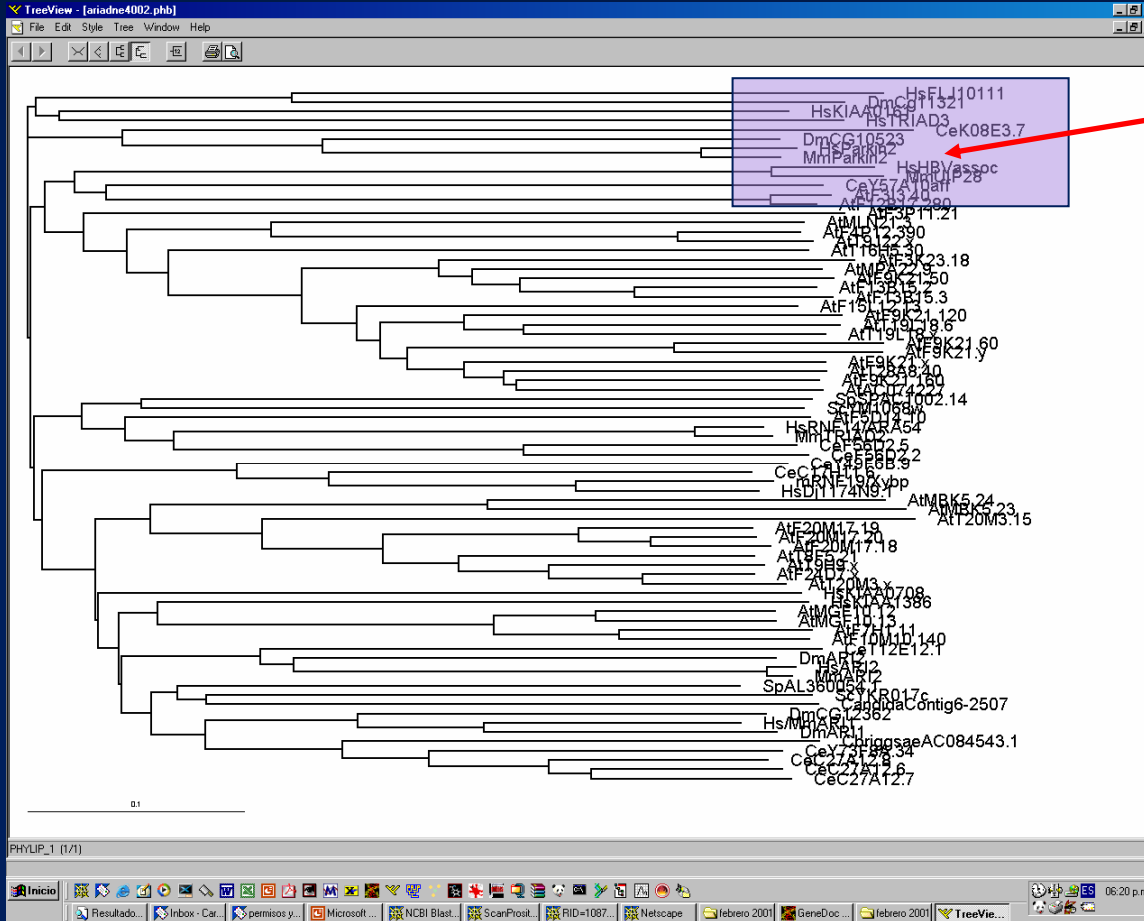
H C C ...etc

El patrón de *aminoácidos conservados evolutivamente* muestra que IBR es parte de un supradominio



Esta característica estructura sólo existe en estas proteínas: **FAMILIA RBR**

Comparaciones y árboles: Parkin es metazoo-específica



Hay ortólogos
de Parkin en
vertebrados
Drosophila
C. elegans,
etc.

Posibles organismos modelo (confirmado)

De vuelta a PUBMED, ¿qué es un RING FINGER?: posible implicación en ubiquitinación

RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination

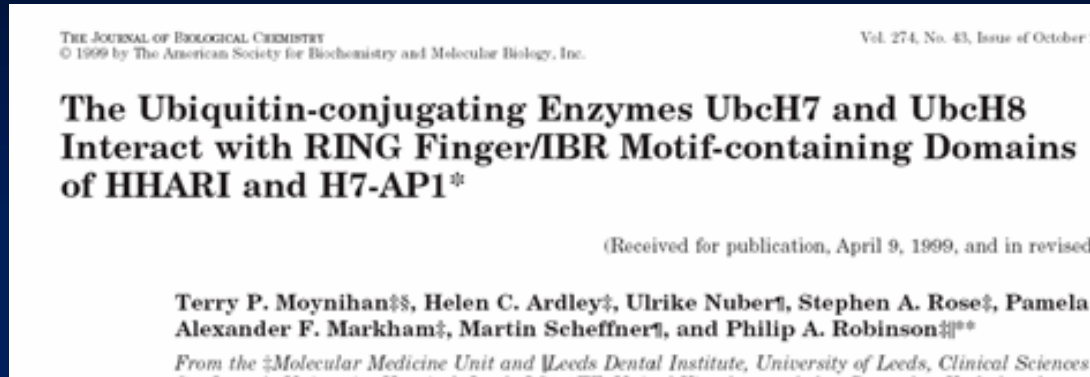
KEVIN L. LORICK^{*}, JANE P. JENSEN^{*}, SHENGYUN FANG, ALBERT M. ONG, SHIGETSUGU HATAKEYAMA
AND ALLAN M. WEISSMAN[‡]

Laboratory of Immune Cell Biology, Division of Basic Sciences, National Cancer Institute, Building 10, Room 1B34, National Institutes of Health, Bethesda, MD 20892-1152

Communicated by Richard D. Klausner, National Institutes of Health, Bethesda, MD, July 26, 1999; accepted for review July 6,

**RING Finger Proteins:
Mediators of Ubiquitin
Ligase Activity**

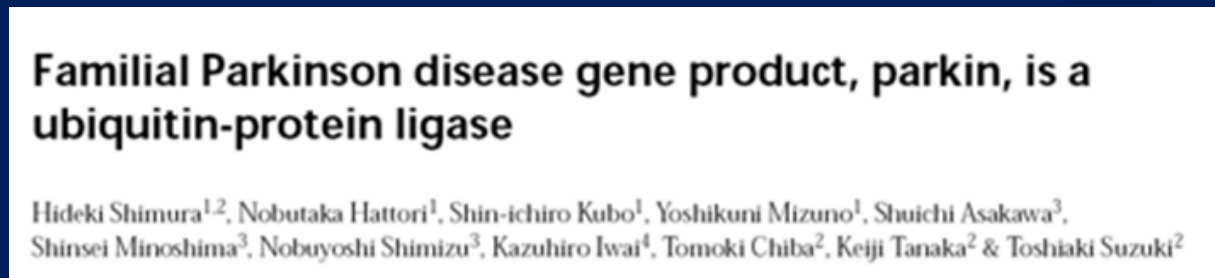
Nuevos datos sobre genes de la familia



HHARI es un parálogo de Parkin y funciona como ubiquitín-ligasa

Posible función bioquímica de Parkin

¡CONFIRMADO!



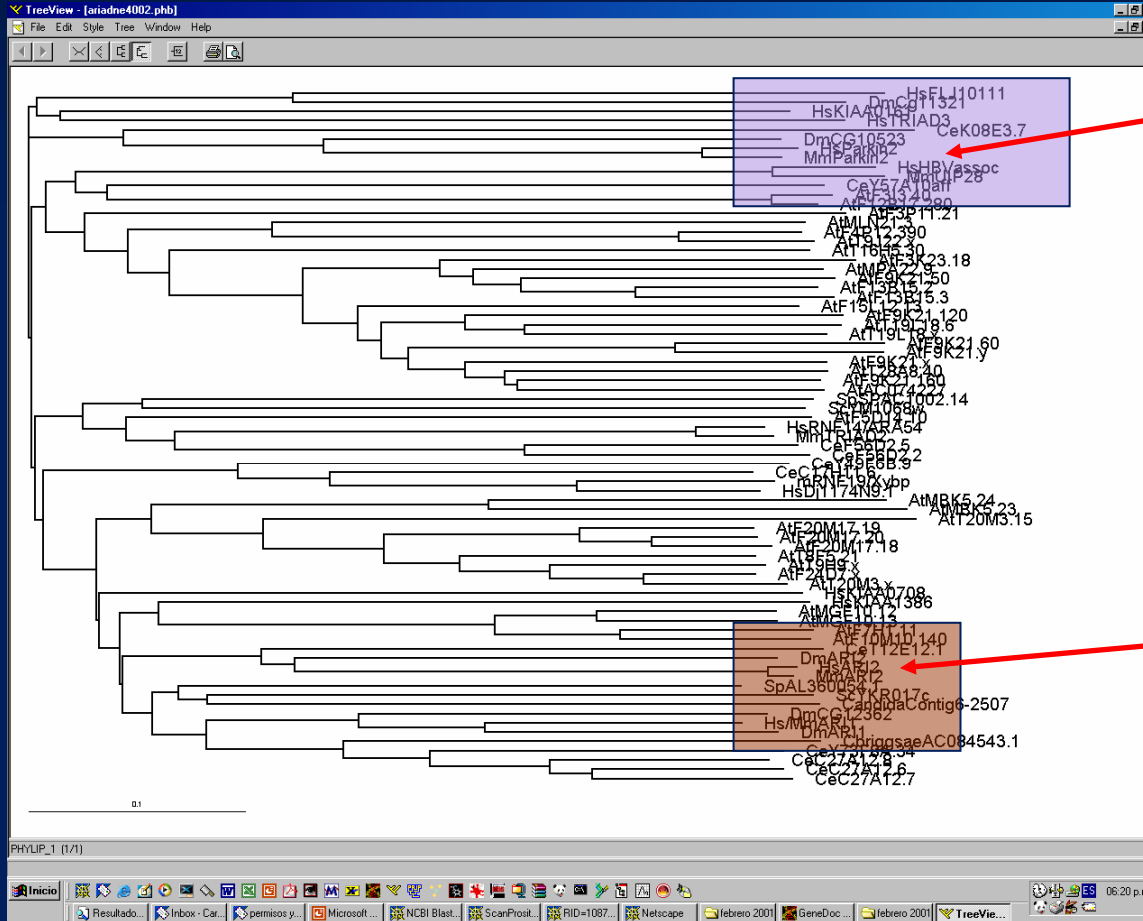
Nuevas cuestiones a resolver

Cuestión 1: ¿pueden ser todas las RBR ubicuitín ligasas?

Cuestión 2: ¿cuáles son sus funciones celulares?

Respuesta cuestión 1:

Probablemente, puesto que Parkin y HHARI están muy alejadas evolutivamente



Parkin

HHARI

Confirmado posteriormente en múltiples trabajos

Cuestión 2: Funciones celulares

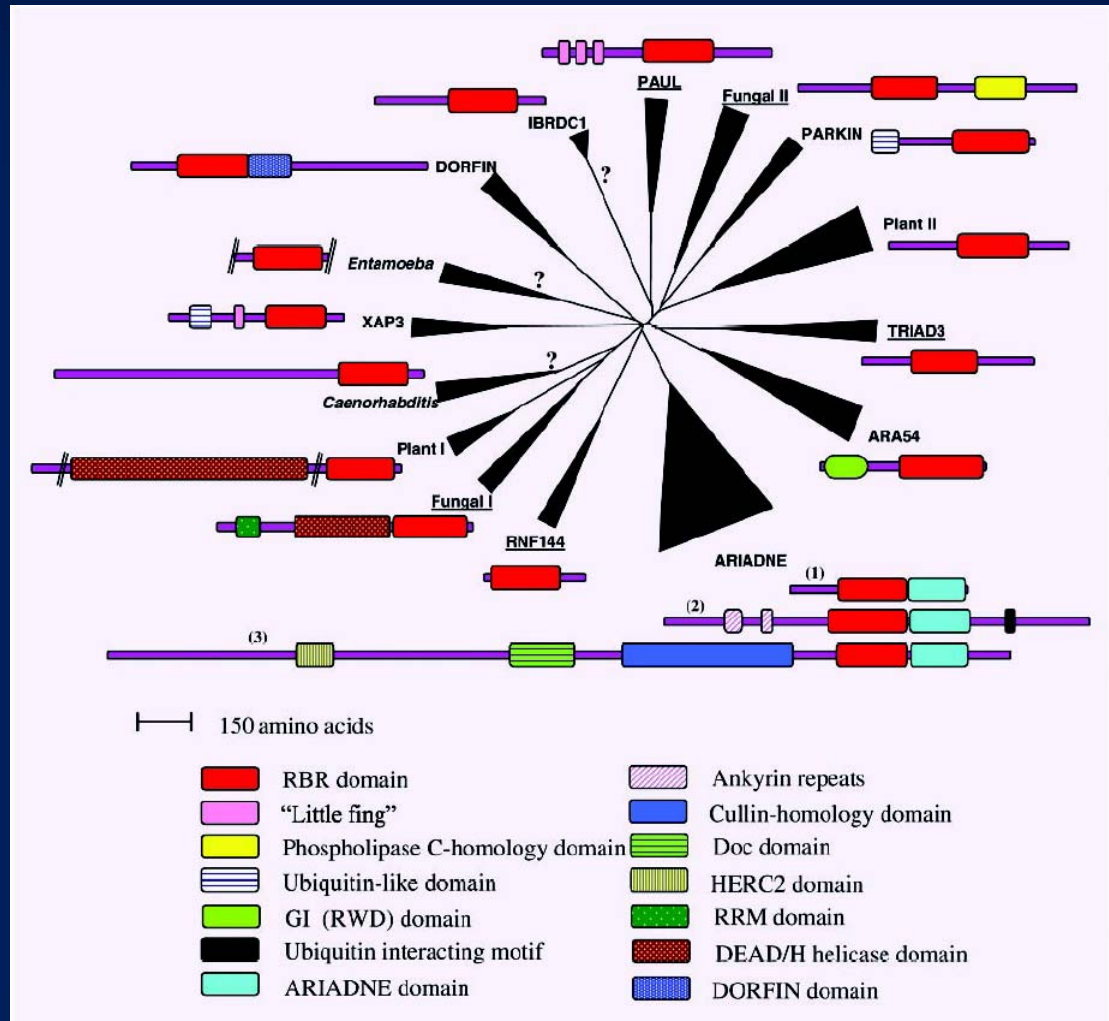
*Hay que profundizar:
nuevos análisis comparativos*

ESTRUCTURAS
RBR

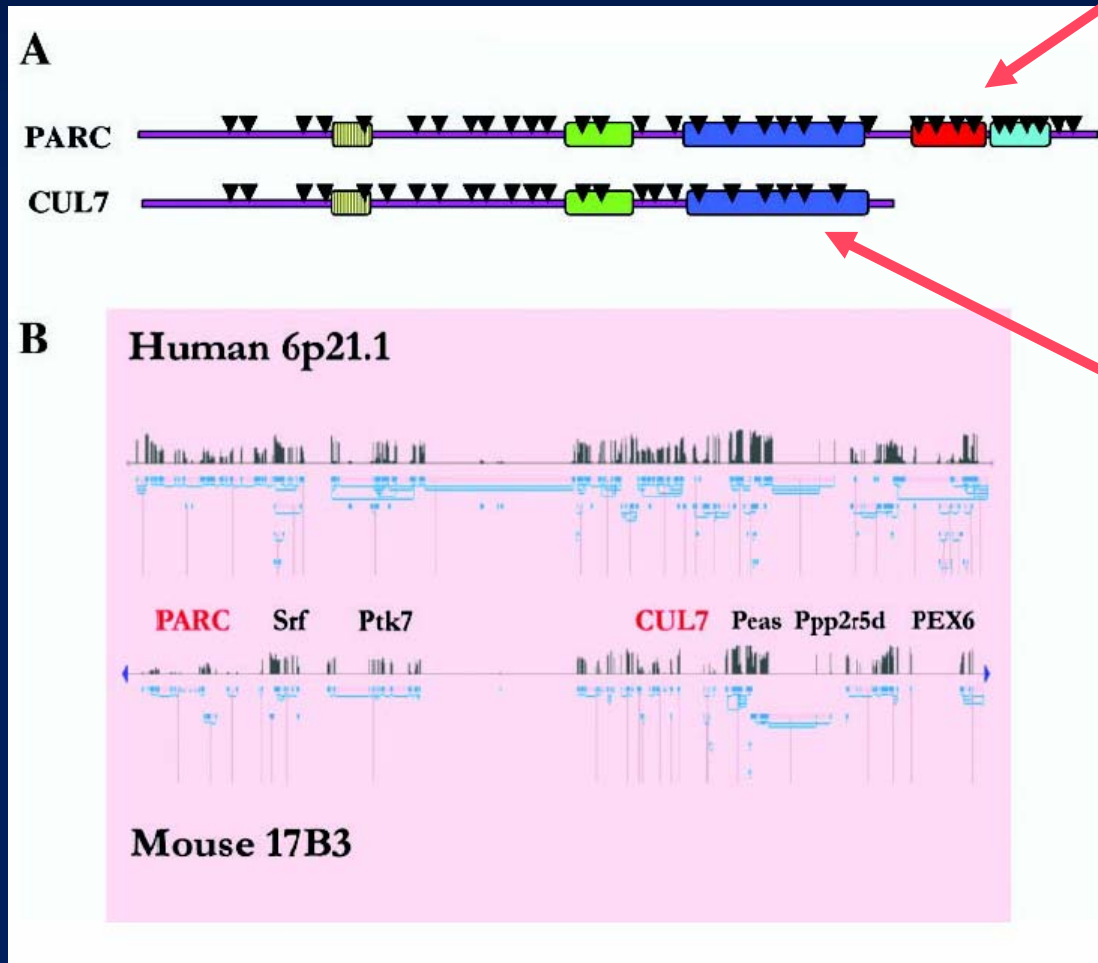
Alto nivel
de variación

Muchos dominios
proteicos distintos

Algunas proteínas
son enormes y
complejas



Sorpresa: genes producto de duplicación + fusión



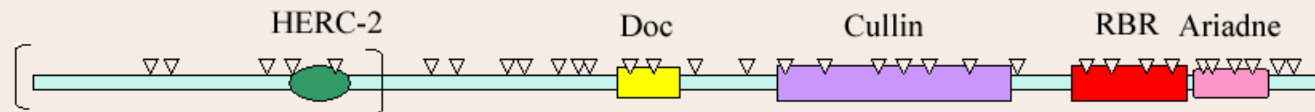
Dominio
RBR

Dominio
Cullin

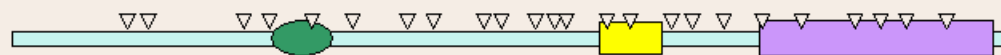


*La estrategia de
la piedra de Rosetta:
genes cuyos productos
interaccionan físicamente
suelen fusionarse*

PARC
(KIAA0708)

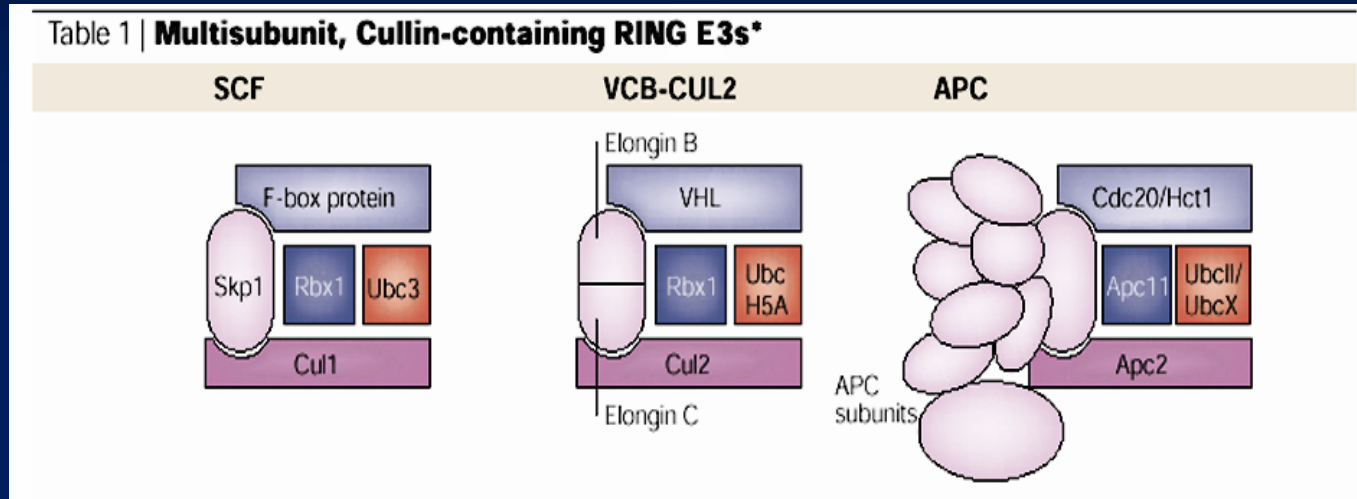


CUL7
(KIAA0076)



100 aa

¿Qué es una Cullin?



FORMAN PARTE DE COMPLEJOS UBICUITÍN-LIGASA JUNTO A PROTEÍNAS CON RING FINGERS

¿Pueden ser las RBRs parte de estos complejos?

Predicción basada en datos comparativos

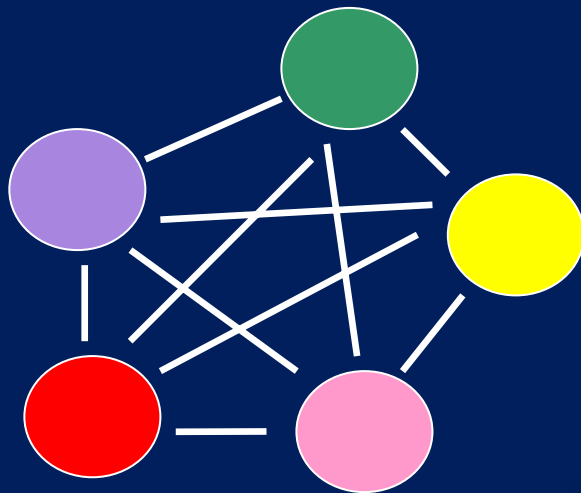
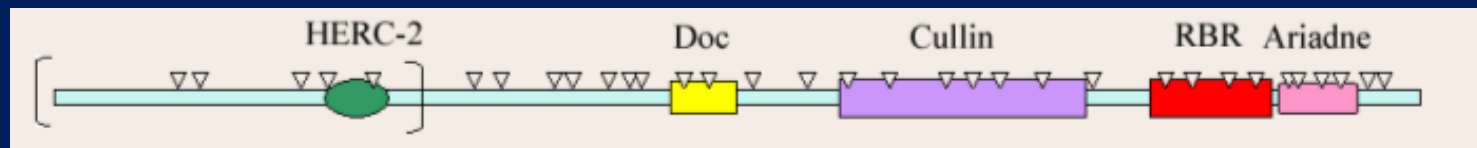
- Las proteínas RBR pueden funcionar como ubiquitín-ligasas en complejos cullin/RING
- **Posteriormente confirmado para Parkin y otras RBRs**

¿Podemos profundizar más?

- Analizar todos los datos en conjunto
- Grafos de dominios proteicos
 - IDEA: Convertir toda la información de dominios en una estructura matemática analizable
 - Dominios = unidades
 - Estar en una misma proteína = enlaces

Generación de un grafo de dominios

- Dominios = unidades
- Estar en una misma proteína = enlaces

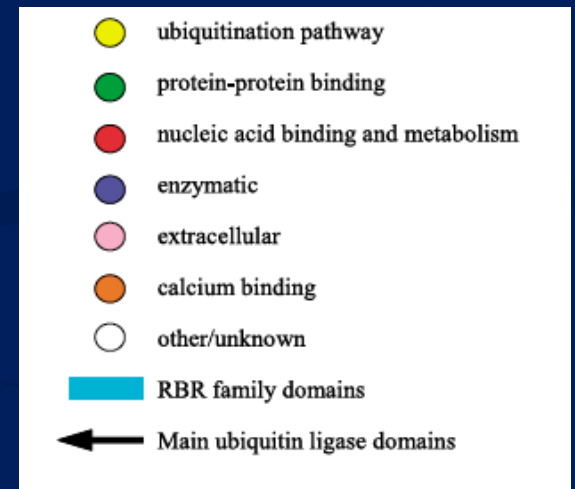
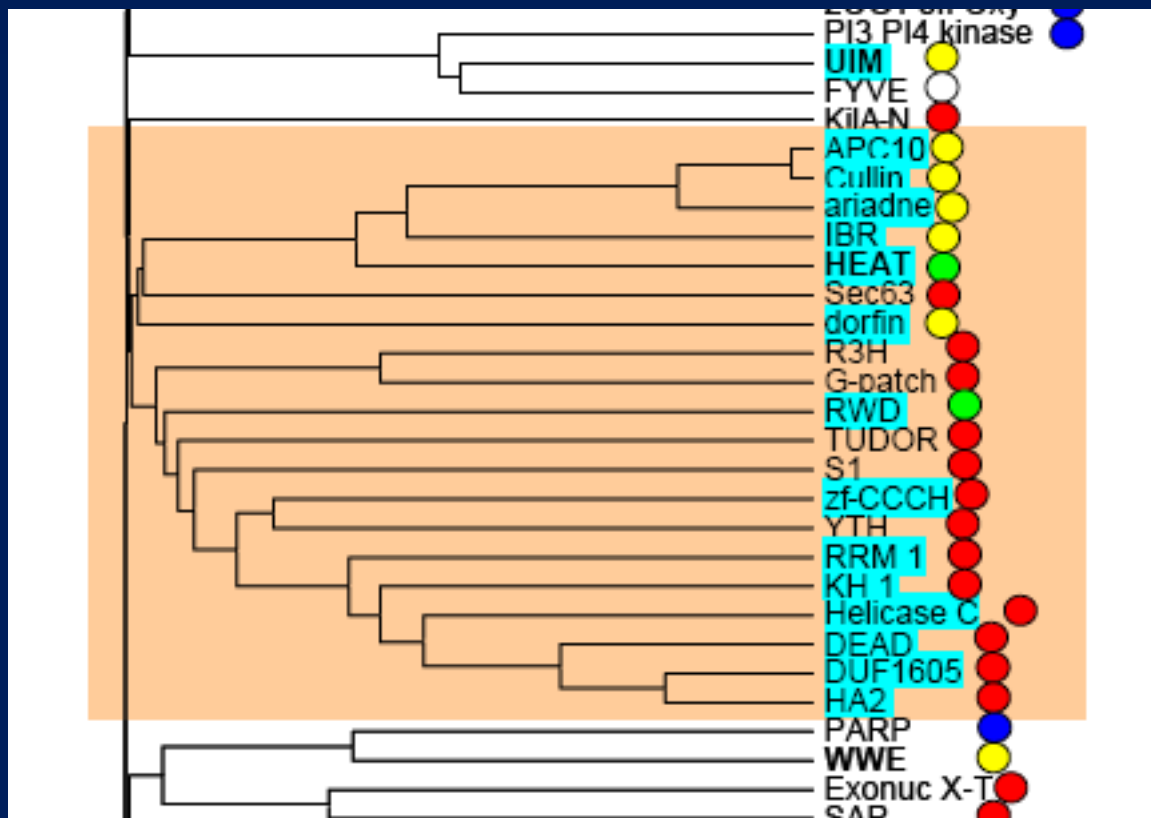


Este grafo contiene toda la información proporcionada por la proteína PARC

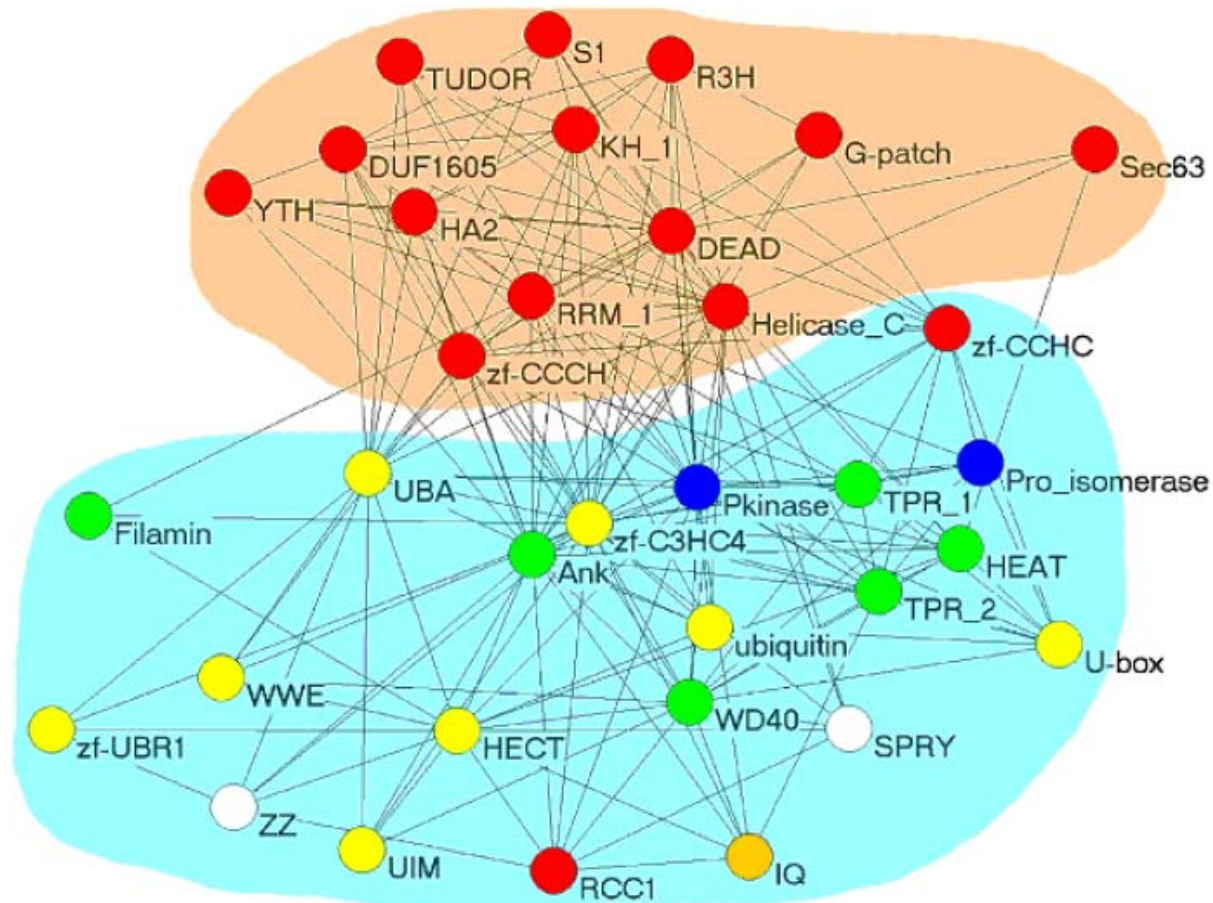
Extrayendo información del grafo total de dominios proteicos

- Pfam: 4516 dominios; 15880 conexiones entre dominios
- 22 dominios distintos en proteínas RBR
- Extraemos todos los dominios que:
 - 1) tienen distancia $d \leq 2$ con los 22 dominios RBR
 - 2) están conectados al menos con dos dominios entre esos 22
- Arbol UVCLUSTER (programa de análisis de grafos)

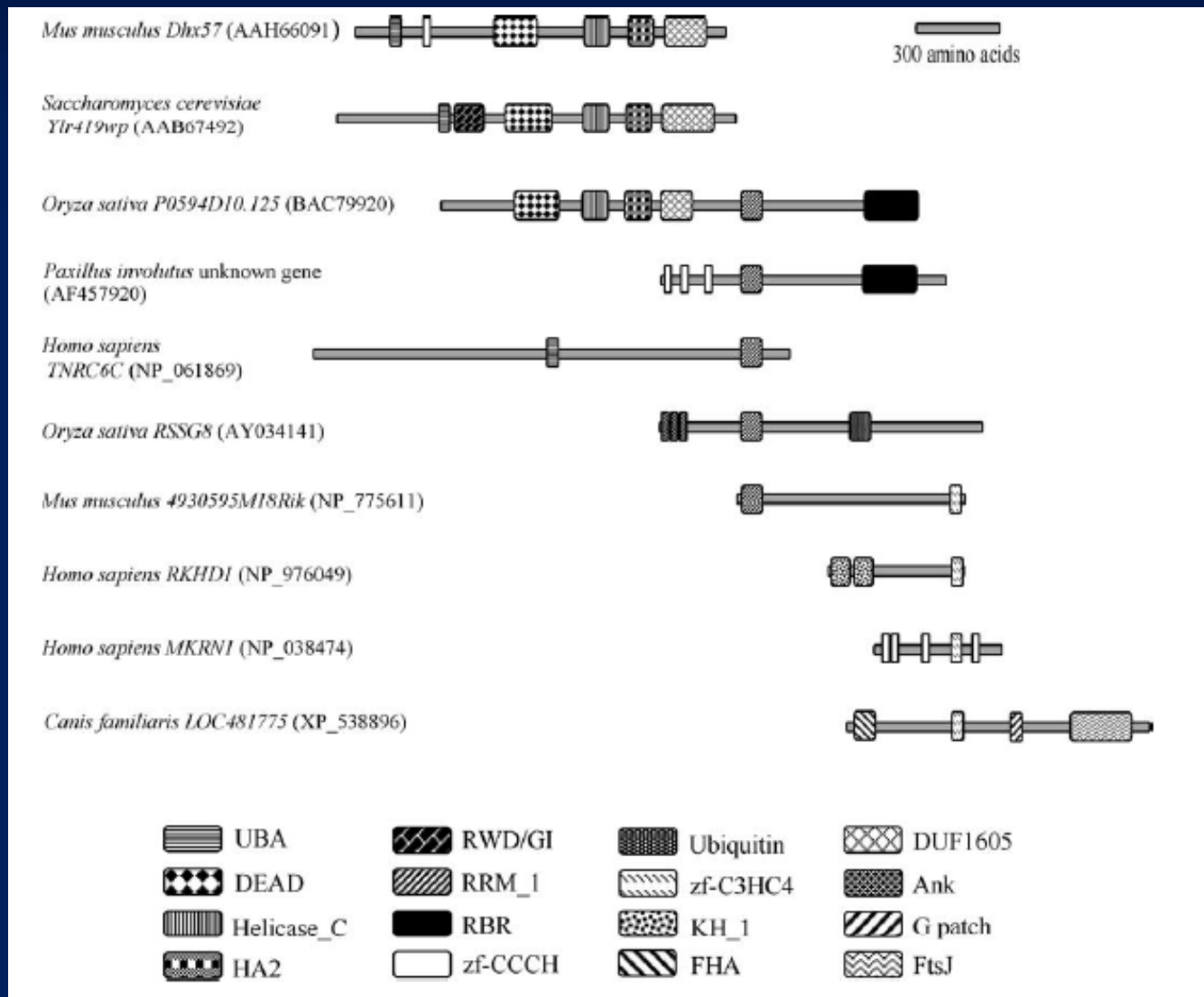
Conexión dominios RBR – metabolismo RNA



Alta conexión dominios ubiquitinación – dominios metabolismo RNA



Múltiples proteínas con dominios asociados a ambas funciones



Es decir...

- Conexión hasta entonces desconocida ubiquitinación – metabolismo RNA
- Hay RBRs particulares implicadas en funciones de regulación del metabolismo del RNA
- ¿Posible función para Parkin/otras RBR?

Conclusiones

- Los datos comparativos ofrecen un marco conceptual en el que entender la investigación funcional
- El análisis de datos proporcionados por múltiples genomas/proteomas permite obtener información imposible de detectar en estudios a menor escala

Agradecimientos

- Parkin

- Ignasi Lucas (Dpto. Genética. Univ. Valencia)
- Alberto Ferrús (Instituto Cajal, CSIC)

- UVCLUSTER:

- Vicente Arnau (Dpto. Informática. Univ. Valencia)