

BIOINFORMÁTICA

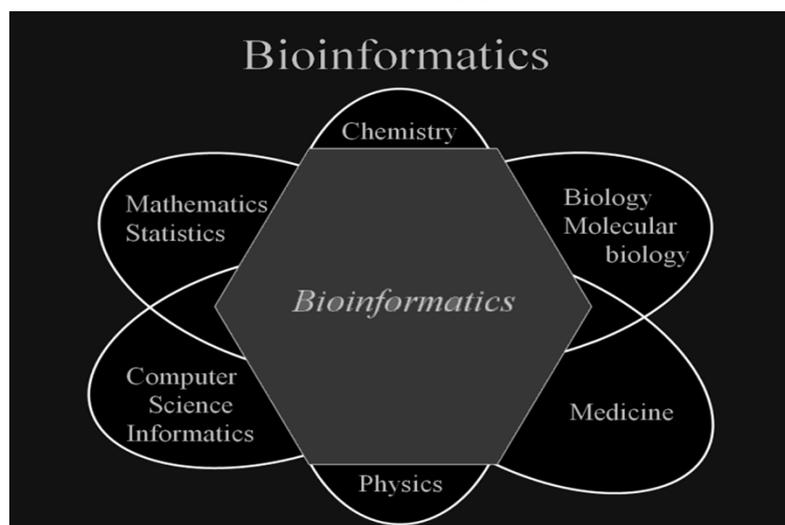
Vicente Arnau Llombart

<http://www.uv.es/~varnau/bioinformatica.pdf>

E-mail: Vicente.Arnau@uv.es

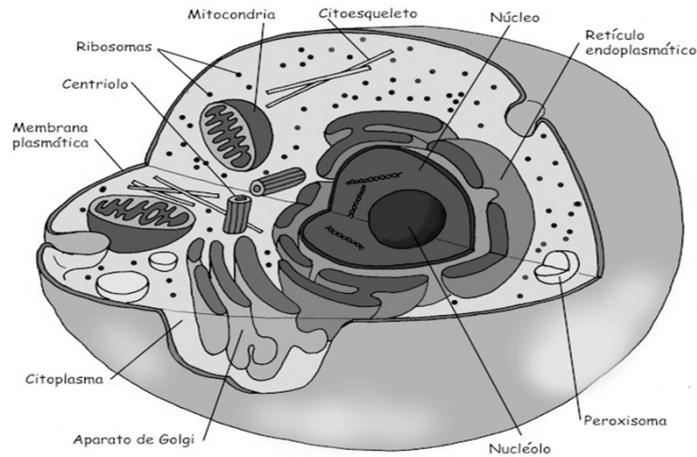
¿Que es la **BIOINFORMÁTICA**?

BIOLOGIA MOLECULAR ↔ INFORMÁTICA



UNIVERSITAT DE VALÈNCIA

La Celula



C. EUCARIOTA Humana posee → 3 Gbases en el NUCLEO y 16 Kbases en las MITOCONDRIAS



El Genoma Humano

La molécula de ADN fue descubierta en 1951 por **James Watson**, **Francis Crick** y **Maurice Wilkins** empleando la técnica de difracción de los rayos X. En 1953, Watson (izquierda) y Francis Crick (derecha) describieron la estructura en doble hélice de la molécula de ADN como una especie de escalera de caracol con muchos escalones. En 1962 ambos recibieron el Premio Nobel de Medicina por su trabajo.

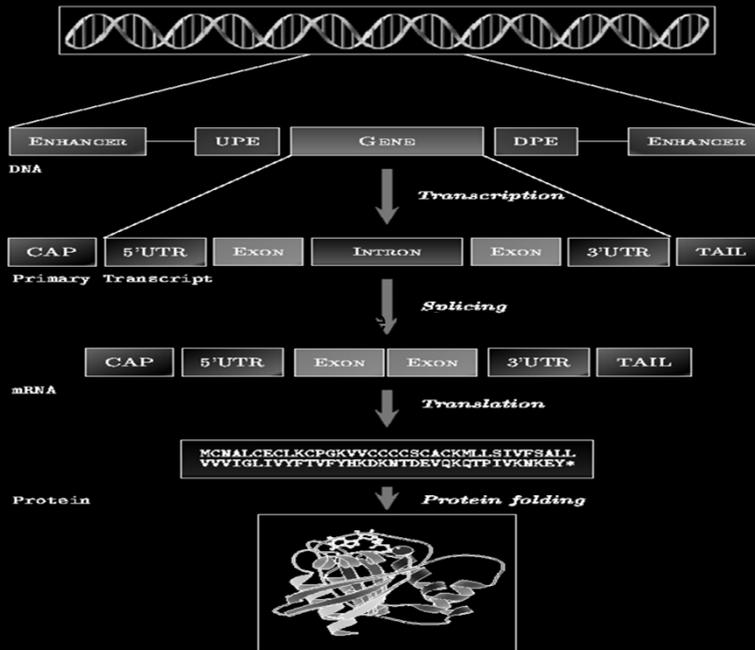


El Genoma Humano

CATGTTTGATTTTTCAATTTTCGTGCGAGATTTTGTGCGTTTTGGGTACCTGGAACATTTCTCGACCA
 ATAGCTCCGGAACATTTAGAAATAACCCATAAAGTGTGCTATCGTTGGAAAGCTCTTTAAATATCTA
 TTTTTTCAGAAAAGATTGTTGTTCTCCGACTAATAGTTTTCGAGCAAATTTGCTGCTAAATGCAAAAATT
 GGAATTTTTATTGGTCAAATTAGCTAATAACTAATTTGTCCAGCAAGAACCACGTGGAGGTTTAGAGC
 ATTCTTGCCAGGAAAAAGGATATAAGGATGAGGTAAACAATGATTACTTCAAATGAAGGGAATAAAGG
 AATTATTGGTTTTGAAATAAAAAAATACAACACAGAAATAAAATGCACCTACCTTTATGGCTCCACGT
 TGGCTACTGTTTAGCAAAAACAATAAAAAGAAATTAACAACACCATTAAAAATCTTAATTTGCATG
 AAAAATTAACAGCGACATAATTTAAAGGTGTAACAAAACAATCAAGAATACTTATTATATTTAAAGGATT
 GCTCTAATTGGTCATGCAGTGTATTATCGTTAGAAGCTCTTTCAATTATCTATTTTTTACAAAAGATT
 ATTGCTCTCCGACTAATAGTTTTTAAATATATAAAAACTATTGGGAAGTTGGCAGTTGTCTCGATTG
 CTAATCGATTTCCCGGATCATTTGCATAGGTAGATTAAATAGTTCACCTTTTTCTAATAGTTTGTCT
 AGTGAATTGAGCGGCTAATTTTTGTTTATACGCTTTTCATTTATTATAATTTATAATTTGAATGTTTT
 GCTTAAAATTGTAATAATCAAGTCTTACATACGGATGTCGGAACATTTATCTAGCAATAACAACGGAA
 CTATTAGAGATAACCGACCGAACATAATCAACTGTGATACAGAACAATGGAATTTTGGTCAAAAATGGATAGTC
 AATCCTCGTTGAGCTGTAATTTGAATCCCATGATTGGCATCAAAATTTGACCTTTAATTTTTACCATTTCGCAAAATTTGATGATTTACCCCTATCGAA
 AATGCAAAAATGGACCCAAAATTTGATTTTTCAAAAATCCTTCAAAAATTTGATTTTTGTCAGTTTTGTTGGTAAAAAGCTGTAAAAAAAGCTATTGGTTCATCTTT
 ATAGCATTTTTGGCCAAAGTTATGAAGAAAAGACCATTGTAATAGCGGATTTTCCAAAAACCGATTTTCGAAATTTCCATAATACATCAATAGGTTTTGGCCA
 TAACITTTGAAATAATAGTCTGAATTCGGAATGTCATATCTGTTGAGCTCGTAATTAATTTCCATTCAAAGTATTCAAAAATGGAATTTCTATTTTACCAATTTT
 TTGCAAAATTTGATGATGTTACCCCTACCGAAAATGCGAAAATGSAACCAAAATTTAAATTTGAAAATCTTTCAAAAAGTGATAGGGATACCTAGCTATTGATAAA
 AAGCTGAAAAAACAGCATTGTTTATCTTATGACCAATTTTAGCAAGTTACGATGTATAGATCTTTGCAAAAAGCGGATTTTCCAAAACCGAATTTTCCA
 TAATAAAAATCTGTTTTTTGGCCATAACTTTAAAAATAAGAGTCTGAATACGGAATGCCATATCTGCTGAACCTGTAATACATTTTCATCAAACTGATTTCAA
 AAATGGAAATCTATTTTTACCCTTTTTGCAAAATTTGATGATGTTACCCCTTACCAGAAATGCGAAAATCGACCCAAAATTTAAATTTTCAAATCTTCAAAAA
 GTGATAGGGACTAGTATTGGTAAAAAGCTGAAAAACAGTCAATGGTTTTATTTATGACCTTTTAGGCAAGTTACGATGATAGATCTTTGCAAAAAGCGGATTTTCCAAA
 AACGATTTTCCATAATAAATATCTGTTTTTGGCCATAACTTTAAAAATAAGAGTCTGAATGGAATGCCATATCTGTTGAGCTCGTAATTAATTTCCATTCAACTGATTTCAA
 AAATGGAAATCTATTTTCCCAATTTTGCAAATTTGATGATGTTACCCCTTACCAGAAATGCGAAAATCGACCCAAAATTTAAATTTAAATCCCTCAAAAAGGATTTGGGATAAATG
 TATTGAAATCTCTTAAAGCTGTTATTTTTAACTATACGATCAITTTAATCCAGTTTAGGCAAGTACGATGATAGATCTTTGCAAAAAGCGGATTTTCCAAAACCGAATTTCCATAATAAATGTTTTTGG
 CCATAGCTTTAAAATAGA.....



The Central Dogma of Molecular Biology



El Genoma Humano

PREGUNTA:

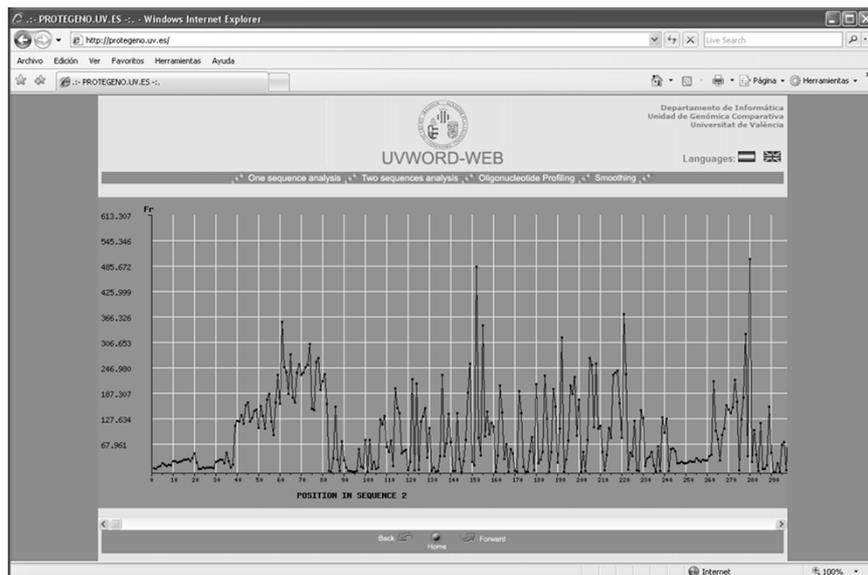
¿Cómo podemos saber que regiones de un gen o de un genoma son regiones codificantes y cuales no?.

RESPUESTA:

```
>Adh
TCAOGTACTGTGGTCGTCCCTTGTTTATGGGCAGGCATCCCTCGTGCCTTGGACTGCTOG
TACATGTGGGGCAGGTTCCTGAAAACGTCGGCATGTTGTCCACTGAGACAACCTGTAAA
CCCGTTCCCGAAACCAGCTGTAAGAGATCOGTAATTGTGTGGCCTGGGGAGACOCCTTCT
CGCTTAGCATOGAAAAGTAACCTCGGGGAAAAGAAAAAATACAAATGTTAAAAATTGTC
TTGTACTTAAGTTGTAAGGTAAGTCTCTCTAATAAGTGGGTCATCTAACATTATAC
ATTTTCATAAAATAATAGATTACAAATGGGTCAAAAATAAATGTTCA)GTGAAAGCTTCCOCTT
CTCAACGTCAATAAAAGCATTFAAAAAAATAGCACAATAATAATFAAAAACTAATTTT
GAAATCTCTTGAACAAGACAGATATTTTGGTTCAGTCGCTGAACAAATCTGTTTACTGT
CTAAAATCTGAAAACCAATTTTCOGACAGCTGACAGCTTCGAAAGAATATAGTACACAA
TTTGCAGTCCAAAAATGAGTACAAAAACAAAACAAATAA)TGAOGAOGOGACTGGGCAT
CTCTTAGTATGAGATATAAGTATTTAAATTTCTAAAAATAAAAGCATTTTTGTCAAAT
AAAATGCAAAAACCGACAAAGTTGATGGAGGGTTTGTAAAAAATAAATTGGAATGTAAA
AGAGTTTCAGTTAGCGCAGGTGGATTTACAGAAAAAATGCAATGCAATTAAACATTAC
ATGTAAGCATGAGTCCATTAAATCATTTCATTTGGTTCAATTOGOGCACGAGCTTAAAT
TATAATGATACAATAAAAAAATGATGATAAAGAGAGACTA . . .
```

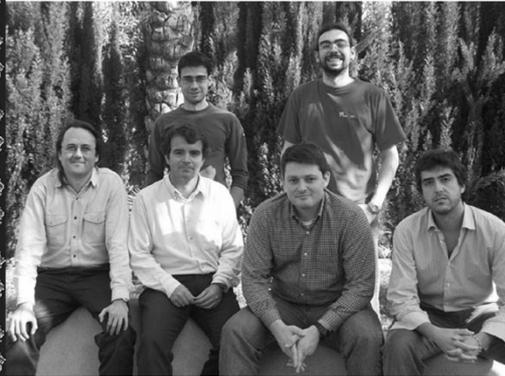


El Genoma Humano



El Genoma Humano

GRUPO DE GENÓMICA DE LA UNIVERSIDAD DE VALENCIA



VNIVERSITAT ID VALÈNCIA
Departamento de Genética



Análisis de Patrones Globales
de Evolución Genómica

TESIS DOCTORAL
Miguel Gallach Caballero
Valencia, 2008



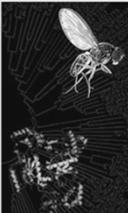
El Genoma Humano

Departamento de Informática
Unidad de Genómica Comparativa
Universitat de València
Instituto de Biomedicina (CSIC)

UVWORD-WEB

Languages:  

[Oligonucleotide Profiling](#) [Oligo Smoothing](#) [Oligo Counting](#) [Singular Oligo Search](#)



UVWORD-WEB is a web tool devised for fast screening of the oligonucleotides composing DNA sequences. The fast analysis of relative abundances of DNA-words (oligonucleotides) is very useful to detect specific oligonucleotides among discrete DNA sequences, chromosomes or even genomes.

UVWORD-WEB implements four kind of analysis based on the exhaustive computing of the oligonucleotides from 1 to 14 nucleotides-length present in any DNA sequence.

Oligonucleotide profiling tool computes for all oligonucleotide frequencies in a DNA sequence (Source). Later, oligos from a second DNA sequence (Target) are scanned to count their frequencies in Source. An additional parameter, window size, must be defined. Window size is the number of consecutive oligonucleotides the user wants to add together to obtain average counts. For window size = 1, no averages are calculated and only the frequencies for every one of the oligonucleotides are computed. For window size > 1, a third column with the average associated variances is added in the table.

In Smoothing option, unlike to the previous option, averages are calculated for sliding windows.

One sequence analysis counts all different oligonucleotides of a given length present in a DNA sequence. The user may choose the option "all" or "between" two defined frequencies (L_INF and L_SUP). The former returns a table in text format with all the oligonucleotides found in the sequence and their respective frequencies. The later returns a table with those oligos which frequencies are equal or higher than L_INF but also equal or lower than L_SUP.

In Two sequences analysis all different oligonucleotides are computed, and the analysis returns a table with all the oligos (and their respective frequencies) that are less or as frequent than a number F_INF in the first sequence and at the same time more or equally frequent than F_SUP in the second sequence.

For any question or suggestions contact with: Vicente.Arnau@uv.es.

<http://protegeno.uv.es/>

Back  Home 



El Genoma Humano

Departamento de Informàtica
Unidad de Genòmica Comparativa
Universitat de València
Instituto de Biomedicina (CSIC)

UWORD-WEB

Languages:

[Oligonucleotide Profiling](#)
 [Oligo Smoothing](#)
 [Oligo Counting](#)
 [Singular Oligo Search](#)

SEQUENCE 1

Homo sapiens

Chromosome 1: **Chromosome 22**

Sequence file 1:

SEQUENCE 2

Homo sapiens

Chromosome 2: **Chromosome 21**

Sequence file 2:

Word size: 12 nucleotides

FREQUENCIES:

Frequency LESS OR EQUAL to **2** in SEQUENCE 1 and GREATER OR EQUAL to **60** in SEQUENCE 2

Use cache

Singular Oligo Search

The idea of this routine is to find out sequences which are enriched in one sequence and rare in another one. The analysis returns a table with all the oligos (and their respective frequencies) which are present at most F_INF times in the first sequence and at least F_SUP times in the second sequence.

[Back](#) [Home](#) [Forward](#)

UNIVERSITAT DE VALÈNCIA

El Genoma Humano

```

=====
= U V W O R D =
= V. Arnau; F. Ferris; I. Marin. 11 - V - 2005 =
=====
35058378 WORDS of 12 nucleotides in File human_chr22
35449345 WORDS of 12 nucleotides in File human_chr21
FINF = 2 FSUP = 60
=====

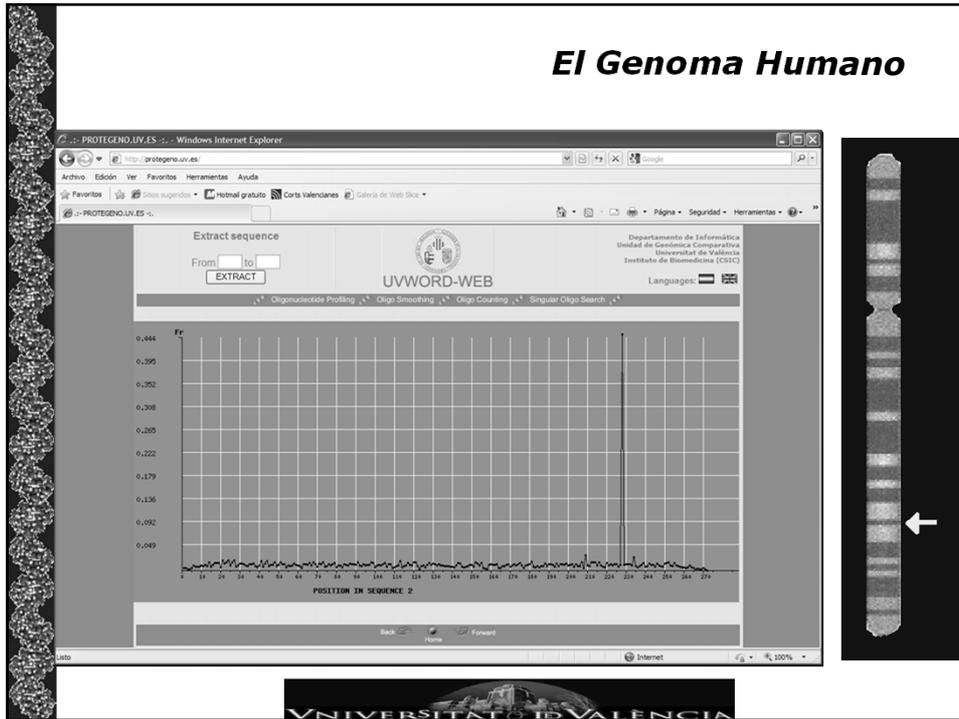
```

WORD	Frec_1	Frec_2
AAATGCCACTT	1	98
AACTCTGGAGT	2	62
AAATGATGGAA	2	61
AAATGCCACTTG	1	82
ATATCCACTTGC	1	72
ATGGAAATGCAAT	2	151
ATGGACTGCAAT	2	82
ATATCCACTTGC	1	80
CAATATGCCACTT	1	79
CAATGCAATGGA	2	190
CATGTTTTGGG	2	75
CCCAITGIIITG	2	60
CTTGAATGCAAT	2	70
GAAGCCACTTCA	0	62
GAATGCAATGTA	2	76
GAATGCAATGGA	1	85
GATGCAATGGA	2	74
GCAATGCAATGG	2	144
GGATGCAATGCG	0	261
GGATGCAATGCG	1	71
GTCCACTTGCAG	2	81
GTGGAAATGCAAT	2	300
GTGGAAATGCAAT	1	114
GTGGAAATGCAAT	1	163
TTGGAAATGCAAT	2	65
TGAGTGGTGAAT	2	142
TGAGTGGTGAAT	1	60
TCCCAATGCAATG	2	155

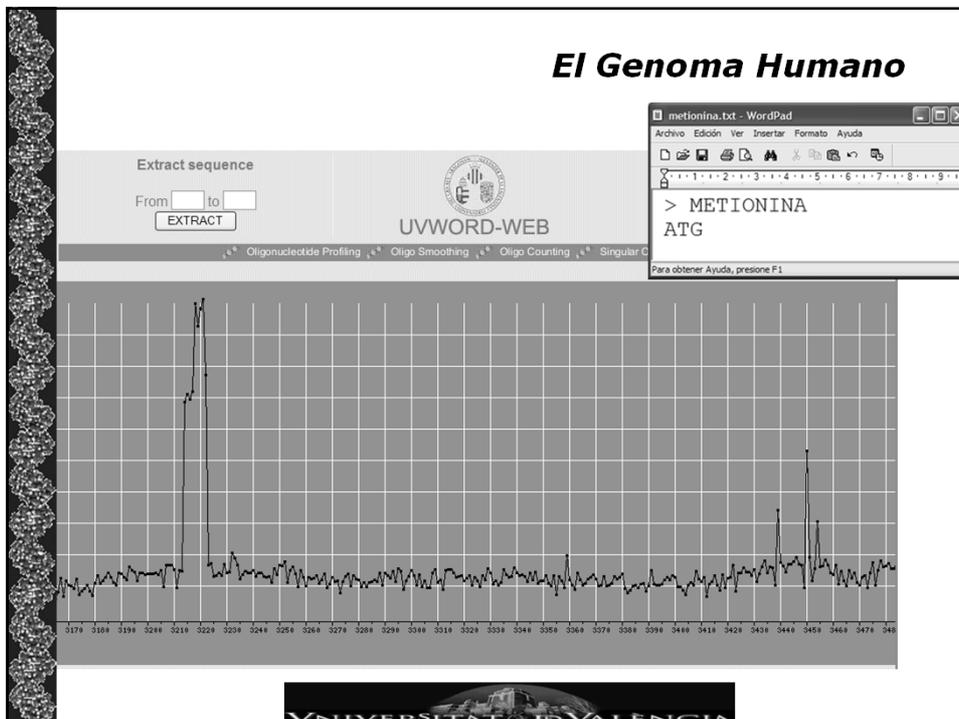
Lite Internet | Modo protegido: activado

UNIVERSITAT DE VALÈNCIA

El Genoma Humano



El Genoma Humano



El Genoma Humano

<http://uvwordweb.uv.es/>

The screenshot shows the UVWORDWEB website interface. On the left is a navigation menu with categories: UVWORD TOOLS (OligoProfile, OligoSmooth, Relative Frequency, FreqWord, SingWord, MultiProfile), GRAPHICAL TOOLS (SeqWord, MaskFreq, Check Gene), and OTHER TOOLS (SMotif, Inverse DNA). The main content area is titled 'OLIGOPROFILE' and contains a 'Select Sequences' section with SOURCE and TARGET dropdowns (both set to '- SPECIES -'), radio buttons for 'Chromosome' and 'Sequence file', and 'Examinar...' buttons. Below this is the 'Length of the oligonucleotide' section with a 'Word size' dropdown set to '1' and 'Nucleotides'. The 'Configuration output' section has a 'Range' input field set to '10000' and an 'Analyze' button. A decorative DNA helix image is at the top right of the main area.

El Genoma Humano

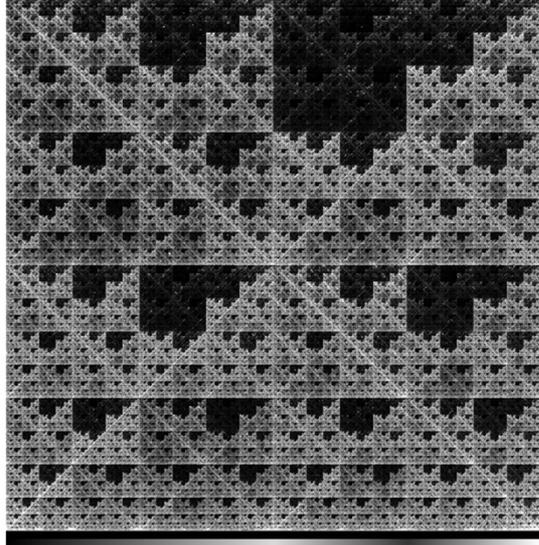
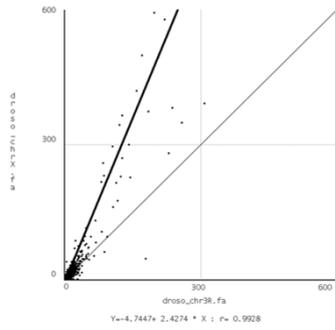
The screenshot shows the UVWORDWEB website interface with the SMOTIF tool selected. The navigation menu is visible on the left. The main content area is titled 'SMOTIF' and contains a 'Select Sequences' section with 'Chromosome' set to 'Drosoph' and 'File sequence to search' selected. Below is the 'Query sequence' section with a 'Write sequence to search' radio button. A browser window is overlaid on the right, showing the URL http://uvwordweb.uv.es/exec/resul_EICV5WI_droso_chrX.txt and the following output:

```
Input file = droso_chrX.fa
Number of nodes created = 65
Levels of the tree = 13

Location_last_base      String
-----
3962087      ACGACGAGTCAC
9018607      ACGACGATTGAC
20047730     ACGACGAGTAC
20972875     ACGACGATTAC
-----

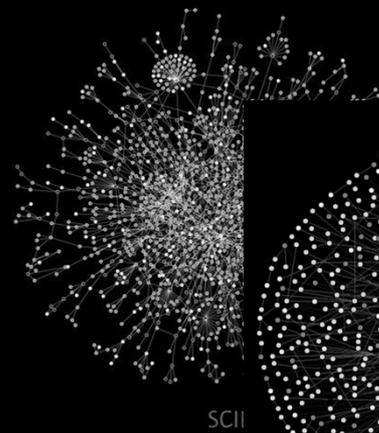
4 strings of 22422827 read bases
```

El Genoma Humano



UNIVERSITAT DE VALÈNCIA

El Proteoma



UVCLUSTER

Colored circles represent proteins (nodes).

- Light blue, known proteins
- Orange, disease proteins
- Yellow, uncharacterized proteins

Interactions (links) are represented by color-coded lines:

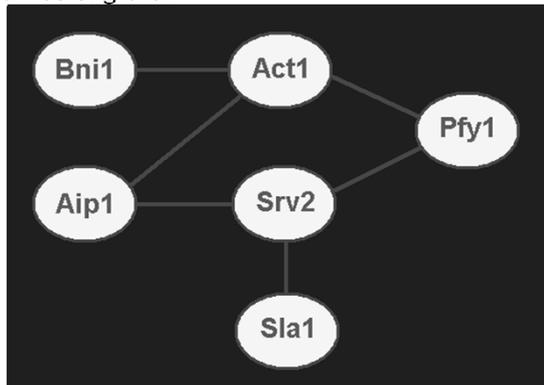
- Red, high confidence (HC) interactions
- Blue, medium confidence (MC) interactions
- Green, low confidence (LC) interactions

UNIVERSITAT DE VALÈNCIA

GRAFO DE INTERACCIONES

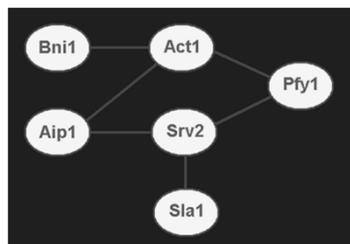
- Leemos de la base de datos un conjunto de interacciones directas entre proteínas y creamos el grafo.

Bni1 ↔ Act1
Aip1 ↔ Srv2
Aip1 ↔ Act1
Pfy1 ↔ Act1
Pfy1 ↔ Srv2
Srv2 ↔ Sla1



UVCLUSTER: ANÁLISIS DE INTERACCIONES ENTRE PROTEÍNAS

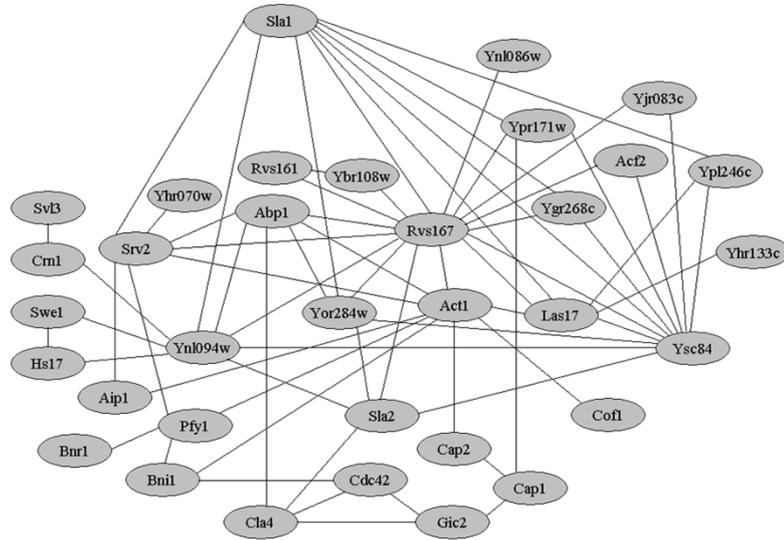
Convertimos Grafo de Interacciones en Tabla de Distancias entre proteínas.



	Bni1	Act1	Aip1	Srv2	Pfy1	Sla1
Bni1	0	1	2	3	2	4
Act1	1	0	1	2	1	3
Aip1	2	1	0	1	2	2
Srv2	3	2	1	0	1	1
Pfy1	2	1	2	1	0	2
Sla1	4	3	2	1	2	0



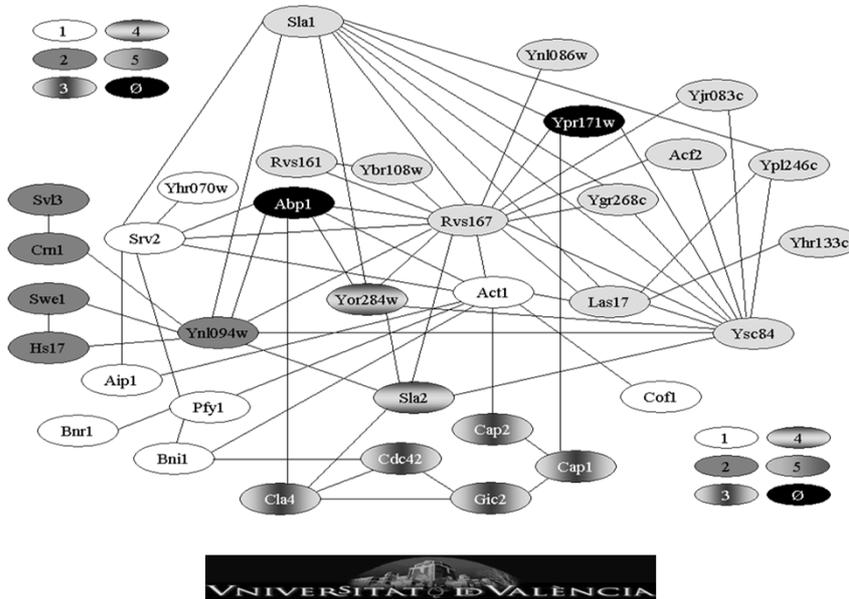
El Proteoma



El Proteoma



El Proteoma



El Genoma Humano

Website screenshot for the Instituto de Biomedicina de Valencia (IBV). The page features a navigation menu with links such as Presentación, Directorio, Biblioteca, Enlaces, Técnicas/Servicios, Memorias, Noticias, Plan Estratégico, and Seminarios. A table lists the research units under three departments: Instituto, Departamento de Genómica y Proteómica, and Departamento de Patología y Terapia Molecular y Celular. A section for the next seminar is also visible, titled 'Próximo Seminario: Apoptosis, inflamación y diferenciación celular, ¿tres en uno?'.

Instituto	Departamento de Genómica y Proteómica	Departamento de Patología y Terapia Molecular y Celular
Presentación	Unidad de Enzimopatología Estructural	Unidad de Biología de la Acción Hormonal
Directorio	Unidad de Proteómica Estructural	Unidad de Biología Vasculard
Biblioteca	Unidad de Genética Molecular	Unidad de Investigaciones Cardíacas
Enlaces	Unidad de Genética y Medicina Molecular	Unidad de Señalización por Nutrientes
Técnicas/Servicios	U. de Cristalografía de Macromoléculas	Unidad de Regeneración Neural
Memorias	U. de Química de Péptidos y Proteínas	Unidad de Patología Metabólica Experimental
Noticias		
Plan Estratégico		
Seminarios	Revistas Electrónicas Contratadas por el CSIC	Intranet CSIC
		Webmail (CSIC)
		Correo (IBV)

BIOINFORMÁTICA



José Enrique Pérez



Joaquín Moreno



El Genoma Humano

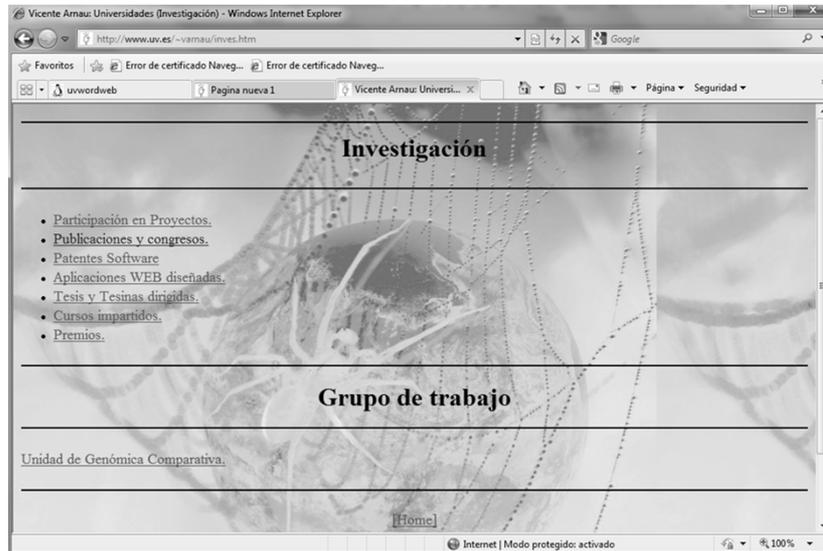
Vicente Arnau Llombart

- [Datos Personales.](#)
- [Docencia.](#)
- [Investigación.](#)
- [Más Personal.](#)

Visitas= 044549



El Genoma Humano



BIOINFORMÁTICA

Vicente Arnau Llombart

<http://www.uv.es/~varnau/bioinformatica.pdf>

E-mail: Vicente.Arnau@uv.es