

## Una proteina nella rete:

### *Introduzione alla bioinformatica*

L'era genomica ha assistito ad una crescita esponenziale delle informazioni biologiche rese disponibili dai progressi nel campo della biologia molecolare. In particolare, il sequenziamento del genoma umano e di altri organismi ha dato un forte impulso a quel settore della bioinformatica che si occupa dello studio del DNA e delle proteine. La grande sfida che la comunità scientifica sta ora affrontando consiste nel cercare di analizzare e capire l'enorme quantità di dati prodotta in laboratorio

La bioinformatica è una nuova disciplina che si occupa dello sviluppo e dell'integrazione delle applicazioni della scienza dell'informazione al servizio della ricerca scientifica in campo biotecnologico. Per fare ciò utilizza strumenti informatici per analizzare i dati biologici che descrivono sequenze di geni, composizione e struttura delle proteine, processi biochimici nelle cellule, etc.

Definizione di BIOINFORMATICA :

(da **Wikipedia**, l'enciclopedia libera) <http://it.wikipedia.org/>

Disciplina ultima arrivata nel campo delle bioscienze, la **bioinformatica** costituisce l'ambizioso tentativo di descrivere dal punto di vista numerico e statistico i fenomeni biologici: storicamente ed epistemologicamente la biologia ha sempre sofferto di una carenza in tal senso rispetto a discipline come la fisica e la chimica, ma oggi la bioinformatica tenta di supplire a questa lacuna fornendo ai risultati tipici della biochimica e della biologia molecolare un corredo di strumenti analitici e numerici davvero promettente.

La bioinformatica principalmente si occupa di:

- \* fornire modelli statistici validi per l'interpretazione dei dati provenienti da esperimenti di biologia molecolare e biochimica al fine di identificare tendenze e leggi numeriche
- \* generare nuovi modelli e strumenti matematici per l'analisi di sequenze di DNA, RNA e proteine la fine di creare un corpus di conoscenze relative alla frequenza di sequenze rilevanti
- \* organizzare le conoscenze acquisite a livello globale su genoma e proteoma in basi di dati al fine di rendere tali dati accessibili a tutti, e ottimizzare gli algoritmi di ricerca dei dati stessi per migliorarne l'accessibilità.

## Banche dati

Una delle attività principali dei bioinformatici consiste nella progettazione, costruzione e uso di **banche dati** di interesse biologico. Una banca dati raccoglie dati e informazioni derivati da esperimenti di laboratorio, da esperimenti *in silico* (cioè utilizzare il dato informatico come punto di partenza per gli esperimenti *in vitro*. Si dice "in silico", in quanto i processori dei calcolatori sono costituiti da silicio) e dalla letteratura scientifica. Le banche dati sono progettate come contenitori costruiti per immagazzinare dati in modo efficiente e razionale al fine di renderli facilmente accessibili a tutti gli utenti: ricercatori, medici, studenti, etc.

Una banca dati è costituita da voci (in inglese *entry*) ciascuna contenente informazioni sull'oggetto caratteristico della banca dati (ad esempio: sequenze nucleotidiche o referenze bibliografiche) insieme a tutte le altre informazioni che si riferiscono a quella entry in particolare).

Una *entry* di una banca dati di sequenze nucleotidiche potrebbe contenere, oltre alla sequenza di una molecola di DNA, il nome dell'organismo cui la sequenza appartiene, la lista degli articoli che riportano dati su quella sequenza, le caratteristiche funzionali (cioè si tratta di un gene o di una sequenza non codificante) e ogni altra informazione ritenuta di interesse.

## Esempio di banca dati: la rubrica telefonica

Ognuno di noi ha esperienza di banche dati come le rubriche telefoniche. In una rubrica telefonica, una *entry* ha come oggetto principale il numero di telefono di uno dei nostri amici o parenti.

La nostra rubrica sarebbe totalmente inutile se insieme ai numeri di telefono non catalogassimo anche il nome e il cognome del possessore del numero di telefono

La nostra rubrica potrebbe essere arricchita anche con altre informazioni: l'indirizzo della persona (di casa e di lavoro), la sua occupazione (se non ci ricordassimo il nome dell'idraulico, dovremmo provare tutti i numeri della rubrica mentre la casa si allaga!!!!), il suo compleanno.

Una banca dati potrà di conseguenza apparire come un elenco di *righe* o come un insieme di *tabelle*

**Bruno Macchi** | dentista | via Calandrino 27 | 02-72597259  
**Carla Cecioni** | autista | piazza Crati 45 | 02-68686868  
**Dante Alighieri** | poeta | via Monti 35 | 02-41563444

NOME	Ercole Palestri
NOME	Dante Alighieri
NOME	Carla Cecioni
NOME	Bruno Macchi
LAVORO	dentista
INDIRIZZO	via Calandrino 27
TELEFONO	02-72597259

## Tipi di banche dati: primarie e specializzate

Le banche dati possono essere di due tipi: primarie o specializzate.

Le **banche dati primarie** contengono informazioni e annotazioni delle sequenze nucleotidiche e proteiche, strutture del DNA e proteine e dati sull' espressione di DNA e proteine.

Le principali banche dati primarie sono: la **EMBL** datalibrary, la **GenBank** e la **DDBJ**. La EMBL datalibrary è la banca dati europea costituita nel 1980 nel laboratorio Europeo di Biologia Molecolare di Heidelberg (Germania). La GenBank è la corrispondente banca americana costituita nel 1982 e la DDBJ è la corrispondente Giapponese. Fra le tre banche dati è stato stipulato un accordo internazionale per cui il contenuto dei dati di sequenza presenti nelle tre banche dati è quasi del tutto coincidente in quanto gli aggiornamenti quotidiani apportati in ciascuna banca dati vengono automaticamente trasmessi alle altre due.

Le **banche dati specializzate** si sono sviluppate successivamente e raccolgono insieme di dati omogenei dal punto di vista tassonomico e/o funzionale disponibili nelle Banche dati Primarie e/o in Letteratura, o derivanti da vari approcci sperimentali, rivisti e annotati con informazioni di valore aggiunto.

Una volta che i dati sono stati archiviati nelle banche dati biologiche è necessario utilizzare alcuni strumenti bioinformatici in modo tale da ricavarne informazioni. Essi si sono sviluppati in base a questi tre processi biologici fondamentali:

- la sequenza del DNA determina la sequenza aminoacidica della proteina (mediante il processo della sintesi proteica);
- la sequenza aminoacidica determina la struttura tridimensionale della proteina;
- la struttura tridimensionale della proteina ne determina la funzione.

La bioinformatica ha focalizzato la sua analisi su dati relativi a questi processi, e di conseguenza le banche dati costituiscono un potente supporto per una vasta gamma di ricerche quali, ad esempio:

- data una sequenza di acidi nucleici o proteica trovare una sequenza simile in banca dati;
- data una struttura proteica trovare, in banca dati, una struttura simile ad essa;
- data una sequenza proteica prevedere una possibile struttura tridimensionale.

I principali strumenti possono essere così organizzati:

### *Ricerca di sequenze simili*

Sequenze omologhe sono sequenze che hanno un gene ancestrale comune. Il grado di similarità fra due sequenze può essere misurato mentre l'omologia è un dato qualitativo.

Esistono una serie di strumenti (es **BLAST**) che possono essere utilizzati per identificare similarità fra nuove sequenze con funzione e struttura sconosciuta e sequenze (archivate nelle banche dati) la cui struttura e funzione sono note.

### *Studio delle funzione delle proteine*

Questo gruppo di programmi (es. **PROSITE**, **SMART**) permette di utilizzare una sequenza per estrarre informazioni su *motif*, domini strutturali dalle banche dati specializzate. Questo potrebbe essere di aiuto per avere informazioni sulla funzione della proteina ignota.

### *Analisi delle strutture*

Questi strumenti permettono di comparare una struttura con una banca dati di strutture note. Molto spesso proteine con struttura simile hanno una stessa funzione, quindi determinare la struttura secondaria/terziaria è cruciale per capire la funzione. (es. **EBI-MSD**)

### *Analisi della sequenza primaria*

Identificare/analizzare l'evoluzione, identificare mutazioni, regioni idrofobiche o altre proprietà che permettano di capire la funzione della proteina. (es. **ENSEMBL**)

## Principali applicazioni della bioinformatica

Numerose possono essere le applicazioni della bioinformatica. Qui citeremo solo un aspetto della **medicina molecolare**. Si ritiene che molte malattie siano associate ad una componente genetica. La malattia, infatti, può essere ereditaria (sono note circa 3000-4000 malattie genetiche come la fibrosi cistica, alcune forme di diabete, etc) oppure essere il risultato di fattori ambientali che causano alterazioni del genoma (tumori, malattie cardiache, ecc). Una branca della bioinformatica studia quali geni siano associati a diverse malattie per capirne più chiaramente le basi molecolari con lo scopo di migliorarne la prevenzione e la cura.

## Una proteina nella rete:

### *Caccia al tesoro bioinformatica*

1. Esiste una proteina umana omologa a Pax6 di zebrafish?

Risposta:

2. Da quali elementi e' possibile capire quale sequenza di quelle ritrovate dal programma sia quella omologa a quella di zebrafish?

Risposta:

3. Prendi nota del codice di identificazione con cui e' indicata la sequenza scelta.

4. Dalla pagina di SwissProt prendi nota del numero TaxID

5. Indica in quale referenza e' possibile trovare informazioni sulla conservazione della sequenza di Pax6 in altri organismi.

6. Indica almeno tre referenze che si riferiscano a malattie associate a mutazioni in Pax6.

7. Indica la funzione della tua proteina.

8. Indica la localizzazione subcellulare della tua proteina.

9. Indica in quali tessuti e' espressa la tua proteina.

10. Indica quale organo colpiscono le malattie associate a mutazioni in questo gene.

11. Prendi nota del numero di accesso della banca dati PDB.

12. Prendi nota di tutti i numeri di accesso della banca dati OMIM.

13. Prendi nota del numero di accesso della banca dati SMART.

14. Qual e' la lunghezza e il peso molecolare di questa proteina?

15. Utilizzando i codici di accesso che hai scritto nelle caselle rosse puoi accedere ad altre informazioni contenute in diverse banche dati.

16. Ora, utilizzando le informazioni che hai raccolto, esegui una ricerca con il motore di ricerca Google utilizzando quattro parole chiave a tua scelta. Prova cambiando le parole e il loro ordine, vince chi riesce a impostare la ricerca in modo che nella prima o nella seconda posizione dei risultati sia presente il link alla pagina di accesso a SwissProt.

## Una proteina nella rete:

### *Caccia al tesoro bioinformatica*

Nel corso di questa attività' utilizzeremo alcune delle piu' importanti banche dati disponibili in rete per cercare informazioni su una proteina. Immaginate di aver appena ottenuto sperimentalmente da tessuti di zebrafish la sequenza della proteina PAX6: seguendo le tracce di questa proteina nella rete scopriremo se abbia un omologo in uomo e troveremo informazioni sulla sua funzione, struttura, localizzazione e sulle malattie associate a mutazioni nella sua sequenza.

1. Apri, con un rapido doppio click, il file sequenza.txt che si trova sul desktop del tuo computer.

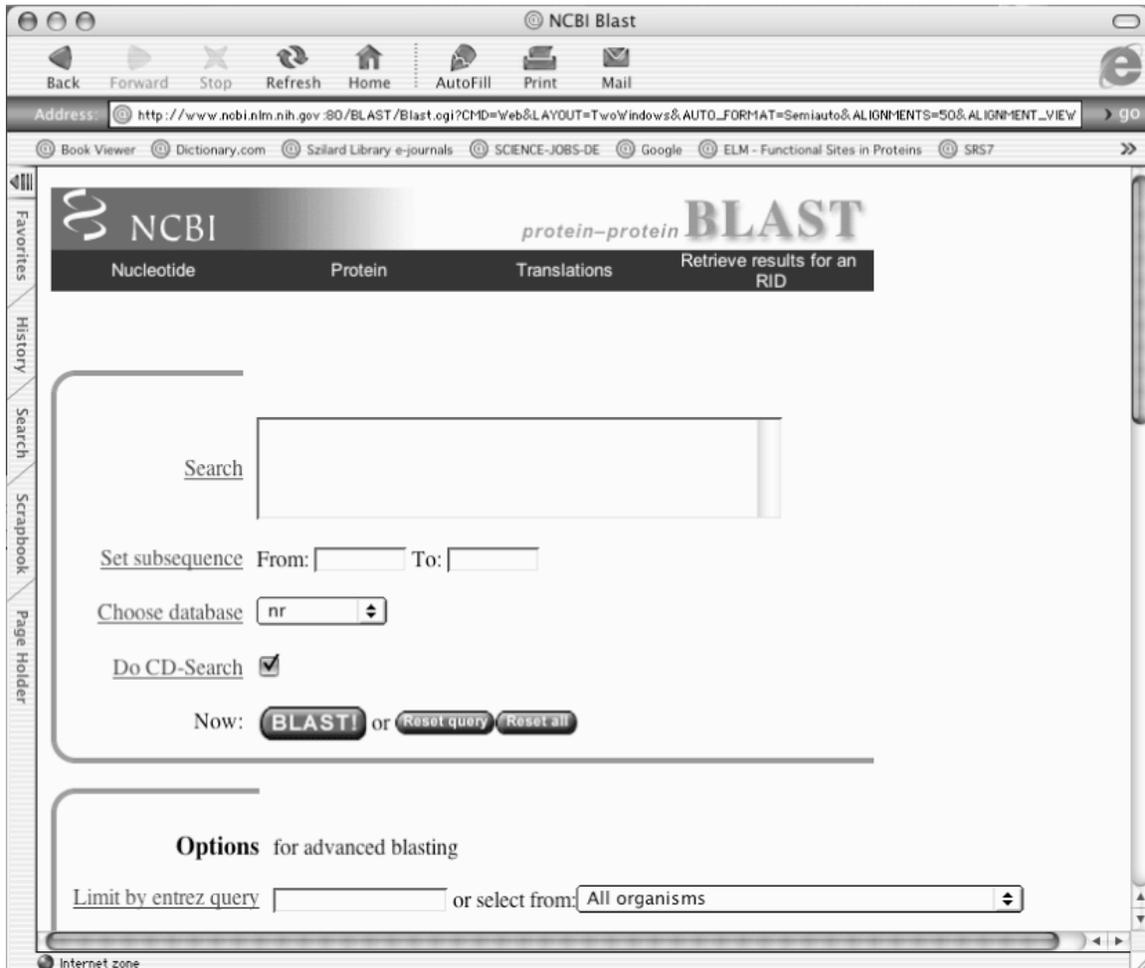
```
MPQKEYYNRATWESGVASMMQNSHSGVNLGGVFNVRPLPDSTRQKIVELAHSGARPCD
ISRILQVSNCGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVGKIAQYKRECPSIFAW
IRDRLLESEGVCTNDNIPSVSSINRVLRLNLASEKQQMGADGMYEKLRLMLNGQTGTWGT
RPGWYPGTSVPGQPNDGDCQQSDGGGENTNSISSNGEDSDETQMLQLKRKLQRNRTSFTQEQ
IEALEKEFERTHYPDVFAERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRQASNS
SSHIPISSSFSTSVYQPIQPQPTTPVSFTSGSMLGRSDTALTNTYSALPPMPSFTMANNLP
MQPSQTSSYSCLPTSPSVNGRSYDITYTPPHMQAHMNSQSMAASGTTSTGLISPGVSPV
QVPGSEPDMSQYWPRLO
```

La sequenza di lettere che vedi rappresenta la sequenza primaria della proteina **Pax6** del pesce zebrafish. Ogni lettera corrisponde ad uno dei venti amminoacidi esistenti, come indicato nella seguente tabella.

Ala	A	Alanina	Leu	L	Leucina
Arg	R	Arginina	Lys	K	Lisina
Asn	N	Asparagina	Met	M	Metionina
Asp	D	Acido aspartico	Phe	F	Fenilalanina
Cis	C	Cisteina	Pro	P	Prolina
Gly	G	Glicina	Ser	S	Serina
Glu	E	Acido glutamico	Thr	T	Treonina
Gln	Q	Glutamina	Try	W	Triptofano
His	H	Istidina	Tyr	Y	Tirosina
Ile	I	Isoleucina	Val	V	Valina

2. Seleziona l'intera sequenza facendo click sul menu' Modifica e scegliendo la voce Seleziona tutto. Quando il testo e' evidenziato con uno sfondo in colore diverso, scegli dal menu' Modifica la voce Copia. Chiudi il file sequenza.txt.

3. Dall'icona del browser sul desktop del tuo computer apri la pagina web di BLAST: <http://www.ncbi.nlm.nih.gov:80/BLAST/> e seleziona l'opzione protein-protein

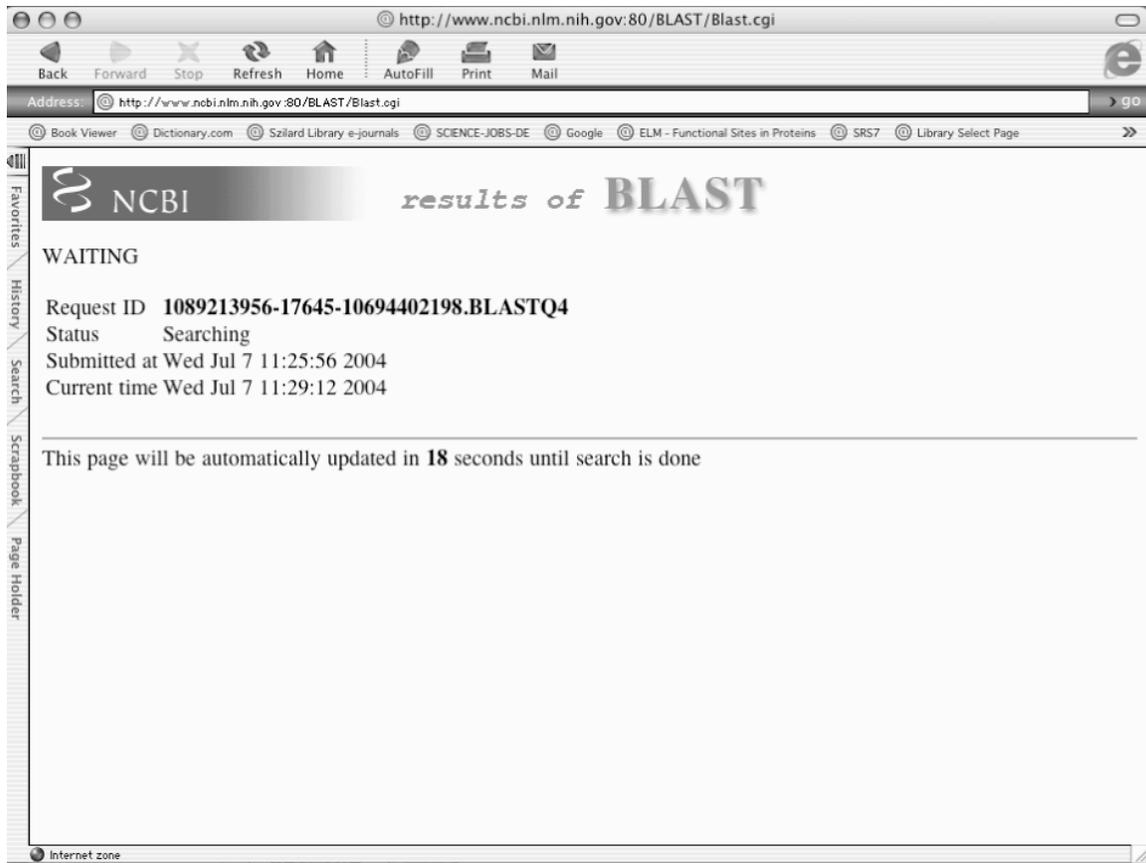


BLAST e' un programma interattivo mantenuto dal sito NCBI (vedi glossario) che serve per confrontare una sequenza (nucleotidica o amminoacidica) con tutte le sequenze registrate nelle banche dati.

4. Con il mouse fai click nella finestra Search e incolla la sequenza che hai copiato precedentemente scegliendo incolla dal menu' Modifica. Seleziona SwissProt nella finestra Choose database. Nella sezione Options nella parte inferiore della pagina, seleziona *Homo sapiens* come organismo. Fai click sul tasto BLAST.

Queste sono le prime informazioni ottenute dal programma sulla tua sequenza, prima di confrontarla con quelle in banca dati. Come puoi vedere la proteina contiene 437 amminoacidi. La rappresentazione grafica evidenzia la presenza di due domini noti nella proteina: il dominio PAX e un omeodominio. I tratti della proteina rappresentati in azzurro (grigio sulla carta) rappresentano le regioni a bassa complessita' (vedi glossario), cioe' regioni caratterizzate da un'inusuale ripetizione di alcuni amminoacidi.

5. Fai click sul tasto Format.



Dovrai aspettare qualche minuto prima di avere il risultato della ricerca perché il programma impiega un certo tempo a confrontare la tua sequenza con quelle disponibili in banca dati.

6. Quando il browser visualizza la pagina dei risultati della ricerca, osservalo da cima a fondo utilizzando la barra di scorrimento laterale.

RID=1089213956-17645-10694402198.BLASTQ4

Back Forward Stop Refresh Home AutoFill Print Mail

Address: http://www.ncbi.nlm.nih.gov:80/BLAST/Blast.cgi

Book Viewer Dictionary.com Szilard Library e-journals SCIENCE-JOBS-DE Google ELM - Functional Sites in Proteins SRS7

**NCBI** results of **BLAST**

**BLASTP 2.2.9 [May-01-2004]**

**Reference:**  
 Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: 1089213956-17645-10694402198.BLASTQ4

**Query=**  
 (437 letters)

**Database:** Non-redundant SwissProt sequences  
 154,183 sequences; 56,416,534 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Taxonomy reports](#)

**Distribution of 102 Blast Hits on the Query Sequence**

Mouse-over to show defline and scores. Click to show alignments

**Color Key for Alignment Scores**

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

1\_17645

Internet zone

Come avrai notato la pagina e' divisa in tre sezioni. La prima presenta lo schema riportato qui sopra. La linea rossa numerata come un righello rappresenta la sequenza che avete inserito per la ricerca e sotto di essa sono rappresentate come spesse linee le varie sequenze di proteine umane trovate in banca dati. Il colore delle rette indica il grado di similarita', il rosso indica il valore piu' elevato come illustrato dalla scala di colore. Come puoi vedere le varie sequenze trovate dal programma hanno diverse lunghezze e diverso grado di similarita'. In questo caso, la prima, pero', mostra quasi la stessa lunghezza e la massima similarita'.

RID=1089213956-17645-10694402198.BLASTQ4

Address: <http://www.ncbi.nlm.nih.gov/80/BLAST/Blast.cgi>

Sequences producing significant alignments:

		Score (bits)	E Value	
<a href="#">gi 6174889 sp P26367 PAX6_HUMAN</a>	Paired box protein Pax-6 (O...	<a href="#">652</a>	0.0	<a href="#">L</a>
<a href="#">gi 3914276 sp O43316 PAX4_HUMAN</a>	Paired box protein Pax-4	<a href="#">236</a>	2e-62	<a href="#">L</a>
<a href="#">gi 1172022 sp P23760 PAX3_HUMAN</a>	Paired box protein Pax-3 (H...	<a href="#">223</a>	8e-59	<a href="#">L</a>
<a href="#">gi 8247951 sp P23759 PAX7_HUMAN</a>	Paired box protein Pax-7 (H...	<a href="#">213</a>	9e-56	<a href="#">L</a>
<a href="#">gi 2506538 sp Q02962 PAX2_HUMAN</a>	Paired box protein Pax-2	<a href="#">212</a>	2e-55	<a href="#">L</a>
<a href="#">gi 548459 sp Q06710 PAX8_HUMAN</a>	Paired box protein Pax-8	<a href="#">203</a>	7e-53	<a href="#">L</a>
<a href="#">gi 417449 sp Q02548 PAX5_HUMAN</a>	Paired box protein Pax-5 (B...	<a href="#">202</a>	2e-52	<a href="#">L</a>
<a href="#">gi 8247950 sp P15863 PAX1_HUMAN</a>	Paired box protein Pax-1 (H...	<a href="#">172</a>	2e-43	<a href="#">L</a>
<a href="#">gi 8247954 sp P55771 PAX9_HUMAN</a>	Paired box protein Pax-9	<a href="#">171</a>	4e-43	<a href="#">L</a>
<a href="#">gi 17380297 sp Q9Y2V3 RX_HUMAN</a>	Retinal homeobox protein Rx ...	<a href="#">88</a>	3e-18	<a href="#">L</a>
<a href="#">gi 13626112 sp Q9H161 ALX4_HUMAN</a>	Homeobox protein aristales...	<a href="#">87</a>	1e-17	<a href="#">L</a>
<a href="#">gi 6174916 sp P54821 PMX1_HUMAN</a>	Paired mesoderm homeobox pr...	<a href="#">86</a>	2e-17	<a href="#">L</a>
<a href="#">gi 27923733 sp Q96Q53 ARX_HUMAN</a>	Homeobox protein ARX (Arist...	<a href="#">84</a>	5e-17	<a href="#">L</a>
<a href="#">gi 25009572 sp Q9NZR4 VSX1_HUMAN</a>	Visual system homeobox 1 (...)	<a href="#">83</a>	1e-16	<a href="#">L</a>
<a href="#">gi 3023579 sp Q15699 CRT1_HUMAN</a>	Cartilage homeoprotein 1 (C...	<a href="#">83</a>	1e-16	<a href="#">L</a>
<a href="#">gi 13626107 sp Q95076 ALX3_HUMAN</a>	Homeobox protein aristales...	<a href="#">82</a>	2e-16	<a href="#">L</a>
<a href="#">gi 17374365 sp P58304 CX10_HUMAN</a>	Homeobox protein CHX10 (Ce...	<a href="#">80</a>	1e-15	<a href="#">L</a>
<a href="#">gi 12644475 sp Q99811 PMX2_HUMAN</a>	Paired mesoderm homeobox p...	<a href="#">78</a>	3e-15	<a href="#">L</a>
<a href="#">gi 8134644 sp Q99453 PMXB_HUMAN</a>	Paired mesoderm homeobox pr...	<a href="#">78</a>	3e-15	<a href="#">L</a>
<a href="#">gi 8134640 sp O14813 PMXA_HUMAN</a>	Paired mesoderm homeobox pr...	<a href="#">78</a>	4e-15	<a href="#">L</a>
<a href="#">gi 6831676 sp O15266 SHOX_HUMAN</a>	Short stature homeobox prot...	<a href="#">77</a>	7e-15	<a href="#">L</a>
<a href="#">gi 48428084 sp Q8N693 ESX1_HUMAN</a>	Extraembryonic, spermatoge...	<a href="#">75</a>	3e-14	<a href="#">L</a>
<a href="#">gi 18202099 sp O60902 SHX2_HUMAN</a>	Short stature homeobox pro...	<a href="#">74</a>	5e-14	<a href="#">L</a>
<a href="#">gi 7531125 sp P56915 GSC_HUMAN</a>	Homeobox protein goosecoid	<a href="#">74</a>	6e-14	<a href="#">L</a>
<a href="#">gi 417427 sp P32243 OTX2_HUMAN</a>	Homeobox protein OTX2	<a href="#">73</a>	1e-13	<a href="#">L</a>
<a href="#">gi 417425 sp P32242 OTX1_HUMAN</a>	Homeobox protein OTX1	<a href="#">71</a>	7e-13	<a href="#">L</a>
<a href="#">gi 6174907 sp Q99697 PIX2_HUMAN</a>	Pituitary homeobox 2 (RIEG ...)	<a href="#">70</a>	8e-13	<a href="#">L</a>
<a href="#">gi 3287764 sp O43186 CRX_HUMAN</a>	Cone-rod homeobox protein	<a href="#">69</a>	2e-12	<a href="#">L</a>
<a href="#">gi 6685480 sp O15499 GSCL_HUMAN</a>	Homeobox protein goosecoid-...	<a href="#">67</a>	7e-12	<a href="#">L</a>
<a href="#">gi 6093723 sp O75364 PIX3_HUMAN</a>	Pituitary homeobox 3 (Homeo...	<a href="#">67</a>	1e-11	<a href="#">L</a>
<a href="#">gi 6093787 sp O75360 PRH1_HUMAN</a>	Homeobox protein prophet of...	<a href="#">65</a>	4e-11	<a href="#">L</a>

Internet zone

Scendendo lungo la pagina con la barra di scorrimento laterale raggiungete la seconda sezione dove sono elencati in blu i numeri di accesso delle varie sequenze trovate in diverse banche dati. A fianco dei numeri di accesso trovate la descrizione breve della proteina a cui si riferiscono e due numeri (Score e E-Value) che esprimono il grado di similarita'. Il piu' utile e' l'E-Value (vedi Glossario), quanto piu' questo punteggio si avvicina a zero tanto piu' elevato e' il grado di similarita'.

**Alignments**

[gi|6174889|sp|P26367|PAX6\\_HUMAN](#)  Paired box protein Pax-6 (Oculorhombin) (Aniridia, type II prot)  
 Length = 422

Score = 652 bits (1682), Expect = 0.0  
 Identities = 358/422 (84%), Positives = 362/422 (85%), Gaps = 4/422 (0%)

Query: 20 MQNSHSGVNQLGGVFNVRPLDPSTRQKIVELAHSGARPCDISRILQVSNCGCVSKILGRY 79  
 MQNSHSGVNQLGGVFNVRPLDPSTRQKIVELAHSGARPCDISRILQVSNCGCVSKILGRY  
 Sbjct: 1 MQNSHSGVNQLGGVFNVRPLDPSTRQKIVELAHSGARPCDISRILQVSNCGCVSKILGRY 60

Query: 80 YETGSIRPRAIGGSKPRVATPEVVGKIAQYKRECPSIFAWAIRDRLLSEGVCTNDNIPSV 139  
 YETGSIRPRAIGGSKPRVATPEVV KIAQYKRECPSIFAWAIRDRLLSEGVCTNDNIPSV  
 Sbjct: 61 YETGSIRPRAIGGSKPRVATPEVVS KIAQYKRECPSIFAWAIRDRLLSEGVCTNDNIPSV 120

Query: 140 SSINRVLRLNLAASEKQQMGADGMYEKLRLMLNGQTGTWTRPGWYPGTSPVGGPNQDGCQQS 199  
 SSINRVLRLNLAASEKQQMGADGMY+KLRMLNGQTG+WGTRPGWYPGTSPVGGP QDGCQQ  
 Sbjct: 121 SSINRVLRLNLAASEKQQMGADGMYDKLRMLNGQTGSWGTRPGWYPGTSPVGGPPTQDGCQQQ 180

Query: 200 DGGGENTNSISSNGEDSDETMXXXXXXXXXXNRTSFTQEIEALEKEFERTHYPDV FAR 259  
 +GGGENTNSISSNGEDSDE QM NRTSFTQEIEALEKEFERTHYPDV FAR  
 Sbjct: 181 EGGGENTNSISSNGEDSDEAQMRLQLKRKLQRNRTSFTQEIEALEKEFERTHYPDV FAR 240

Query: 260 ERLAAKIDLPEARIQVWFSNRRAKWRREEKLXXXXXXXXXXXXXXXXXXXXXXXXXVYQPIP 319  
 ERLAAKIDLPEARIQVWFSNRRAKWRREEKL VYQPIP  
 Sbjct: 241 ERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRQASNTPSHIPSSSFSTSVYQPIP 300

Query: 320 QPTTPV-SFTSGSMLGRSDTALTNTYSALPPMPSFTMANNLPMQ---PSQTSSYSCLMPT 375  
 QPTTPV SFTSGSMLGR+DTALTNTYSALPPMPSFTMANNLPMQ PSQTSSYSCLMPT  
 Sbjct: 301 QPTTPVSSFTSGSMLGRSDTALTNTYSALPPMPSFTMANNLPMQPPVPSQTSSYSCLMPT 360

Query: 376 SPSVNGRSYDITYTPPHMQAHMNSQMAASGTTSTGLIXXXXXXXXXXXXXXXXXDMSQYWPR 435  
 SPSVNGRSYDITYTPPHMQ HMNSQ M SGTTSTGLI DMSQYWPR  
 Sbjct: 361 SPSVNGRSYDITYTPPHMQTHMNSQPMGTSGTTSTGLISPGVSVPPVQVPGSEPDMSQYWPR 420

Query: 436 LQ 437  
 LQ  
 Sbjct: 421 LQ 422

[gi|3914276|sp|O43316|PAX4\\_HUMAN](#)  Paired box protein Pax-4  
 Length = 350

Score = 236 bits (601), Expect = 2e-62  
 Identities = 140/276 (50%), Positives = 167/276 (60%), Gaps = 49/276 (17%)

La terza sezione di questa pagina mostra l'allineamento amminoacido per amminoacido fra la sequenza da noi immessa (indicata con *Query*) e ciascuna delle sequenze trovate dal programma (indicata con *Sbjct*). In mezzo a queste due e' inserita una sequenza che comprende solo gli amminoacidi comuni ad entrambe. Dove la nostra sequenza (*Query*) presenta delle X anziche' gli amminoacidi che occupano realmente quella posizione, significa che il programma non ne ha tenuto conto (si dice che li ha *mascherati*) perche' sono regioni a bassa complessita' e quindi non statisticamente significative.

7. Prendi nota del simbolo della proteina che corrisponde alla prima sequenza: Pax6\_human. Come ricorderete dalla rappresentazione grafica iniziale, la prima e' la sequenza con la maggior similarita'. Questo significa che, con ogni probabilita', la proteina umana omologa a Pax6 di zebrafish si chiama, anche in questo organismo, Pax6. Con questo dato siete pronti per proseguire la caccia al tesoro nella prossima banca dati: SwissProt.

8. Apri il sito della banca dati SwissProt all'indirizzo:  
<http://www.expasy.org/sprot/sprot-top.html>.

The screenshot shows a web browser window titled "ExPASy - Swiss-Prot and TrEMBL". The address bar contains "http://www.expasy.org/sprot/sprot-top.html". The page features a navigation menu with links: "ExPASy Home page", "Site Map", "Search ExPASy", "Contact us", "PROSITE", and "Proteomics tools". A search bar is present with the text "Swiss-Prot/TrEMBL" and "Go" and "Clear" buttons. The main content area includes the "swissprot" logo and the text "Swiss-Prot Protein knowledgebase TrEMBL Computer-annotated supplement to Swiss-Prot". Below this, it states "The UniProt Knowledgebase consists of:" followed by a bulleted list: "Swiss-Prot; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [More details / References / Linking to Swiss-Prot / User manual / Recent changes / Commercial users / Disclaimer].", and "TrEMBL; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot." It also mentions "These databases are developed by the Swiss-Prot groups at SIB and at EBI." A section titled "UniProt Release 2.0 consists of:" lists "Swiss-Prot Release 44.0 of 05-Jul-2004: 153871 entries (More statistics)" and "TrEMBL Release 27.0 of 05-Jul-2004: 1333917 entries (More statistics)". A grey box contains "> Swiss-Prot headlines" and "New full release is available (44.0) (Read more...)". At the bottom, a section titled "Access to Swiss-Prot and TrEMBL" lists: "SRS - Access to Swiss-Prot, TrEMBL and other databases using the Sequence Retrieval System", "Full text search in Swiss-Prot and TrEMBL", "Advanced search in Swiss-Prot and TrEMBL by description, gene name and organism (can be used to create html links to Swiss-Prot/TrEMBL queries)", and "Taxonomy browser (NEWT)".

9. Inserisci nella finestra Search il codice d'identificazione della nostra proteina (Pax6\_human) e fai click sul tasto Go.

10. La pagina dei risultati e' molto lunga e divisa in sezioni, ciascuna segnalata con una striscia blu. Con la barra di scorrimento laterale osserva i vari tipi di informazioni disponibili. Poi torna all'inizio.

The screenshot shows a web browser window displaying the 'NiceProt View of Swiss-Prot: P26367' page. The browser's address bar shows the URL: <http://www.expasy.org/cgi-bin/niceprot.pl?P26367>. The page features a navigation menu with links for 'ExPASy Home page', 'Site Map', 'Search ExPASy', 'Contact us', and 'Swiss-Prot'. A search bar contains 'Swiss-Prot/TrEMBL' and 'pax 6'. The main heading is 'NiceProt View of Swiss-Prot: P26367'. Below the heading are several tabs: '[Entry info]', '[Name and origin]', '[References]', '[Comments]', '[Cross-references]', '[Keywords]', '[Features]', '[Sequence]', and '[Tools]'. A note states: 'Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.' The 'Entry information' section is highlighted and contains the following data:

Entry name	PAX6_HUMAN
Primary accession number	P26367
Secondary accession number	Q99413
Entered in Swiss-Prot in	Release 23, August 1992
Sequence was last modified in	Release 38, July 1999
Annotations were last modified in	Release 44, July 2004

The 'Name and origin of the protein' section is also highlighted and contains the following data:

Protein name	Paired box protein Pax-6
Synonyms	Oculorhombin Aniridia, type II protein
Gene name	Name: PAX6 Synonyms: AN2
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

The 'References' section is highlighted and contains the following information:

[1] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=92103673;PubMed=1684738 [NCBI, ExPASy, EBI, Israel, Japan]  
Tan C C T, Himmelfarb H, Mitsu H, Weil M M, Monaghan P, Jordan T, van Heeringen V, Hastie N D.

At the bottom of the page, a link is provided: [Link: http://www.expasy.org/cgi-bin/get-entries?RA=Hastie N.D.](http://www.expasy.org/cgi-bin/get-entries?RA=Hastie N.D.)

In quest'area della pagina puoi trovare informazioni sul nome della proteina e i suoi sinonimi. Inoltre e' indicato il nome del gene e alcune indicazioni tassonomiche sull'organismo (in questo caso l'uomo).

11. Prendi nota del numero d'identificazione tassonomica TaxID (9606). Esso identifica quest'organismo nella banca dati dell'NCBI che contiene informazioni tassonomiche. Fai scorrere la pagina con la barra di scorrimento laterale fino alla sezione References.

NiceProt View of Swiss-Prot: P26367

Address: <http://www.expasy.org/cgi-bin/niceprot.pl?P26367>

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

### References

[1] SEQUENCE FROM NUCLEIC ACID.  
 MEDLINE=92103673;PubMed=1684738 [NCBI, ExPASy, EBI, Israel, Japan]  
 Ton C.C.T., Hirvonen H., Miwa H., Weil M.M., Monaghan P., Jordan T., van Heyningen V., Hastie N.D., Meijers-Heijboer H., Drechsler M., Royer-Pokora B., Collins F.S., Swaroop A., Strong L.C., Saunders G.F.;  
 "Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region.";  
 Cell 67:1059-1074(1991).

[2] SEQUENCE FROM NUCLEIC ACID.  
 MEDLINE=94258210;PubMed=1345175 [NCBI, ExPASy, EBI, Israel, Japan]  
 Glaser T., Walton D.S., Maas R.L.;  
 "Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene.";  
 Nat. Genet. 2:232-239(1992).

[3] SEQUENCE FROM NUCLEIC ACID (ISOFORM PAX6).  
 TISSUE=Lung;  
 DOI=10.1073/pnas.242603899;MEDLINE=22388257;PubMed=12477932 [NCBI, ExPASy, EBI, Israel, Japan]  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 Proc Natl Acad Sci U S A 99:16899-16903(2002)

Link: <http://www.expasy.org/cgi-bin/get-entries?RA=Dickson M.C.>

In questa sezione puoi trovare elencati i dati di alcuni articoli relativi alla nostra proteina, Pax6.

12. Cerca in quale referenza e' possibile trovare informazioni sulla conservazione della sequenza di Pax6 in altri organismi. (risposta: [2]) Cerca almeno tre referenze che si riferiscano a malattie associate a mutazioni in Pax6. (risposta: dalla [14] in poi). Fai scorrere la pagina fino alla sezione Comments.

NiceProt View of Swiss-Prot: P26367

Address: <http://www.expasy.org/cgi-bin/niceprot.pl?P26367>

Book Viewer Dictionary.com Szilard Library e-journals SCIENCE-JOBS-DE Google ELM - Functional Sites in Proteins SRS7

Ann. N.Y. Acad. Sci. 72:1505-1510(2003).

### Comments

- **FUNCTION:** Transcription factor with important functions in the development of the eye, nose, central nervous system and pancreas. Required for the differentiation of pancreatic islet alpha cells (*By similarity*). Competes with PAX4 in binding to a common element in the glucagon, insulin and somatostatin promoters (*By similarity*). The Pax6-5A isoform appears to function as a molecular switch that specifies target genes.
- **SUBCELLULAR LOCATION:** Nuclear.
- **ALTERNATIVE PRODUCTS:**
  - **Alternative splicing [3 named forms] Display all isoform sequences in Fasta format**  
*Comment:* Experimental confirmation may be lacking for some isoforms.

<b>Name</b>	<b>Pax6</b>
Isoform ID	P26367-1
This is the isoform sequence <u>displayed in this entry</u> .	

<b>Name</b>	<b>Pax6-5A</b>
Isoform ID	P26367-2
Features which should be applied to build the isoform sequence: <u>VSP_002366</u> .	

<b>Name</b>	<b>Pax6-5A,6*</b>
Isoform ID	P26367-3
The sequence of this isoform is not described.	

- **TISSUE SPECIFICITY:** Fetal eye, brain, spinal cord and olfactory epithelium. Isoform Pax6-5A is less abundant than the PAX6 shorter form.
- **DEVELOPMENTAL STAGE:** Expressed in the developing eye and brain.
- **DISEASE:** Defects in PAX6 are the cause of aniridia type II (AN2) [[MIM:106210](#)]. AN2 is a bilateral panocular disorder characterized by complete or partial absence of the iris, absence of the fovea and malformations of the lens and anterior chamber. Severe age-related corneal degeneration is a frequent complication which contributes to a poor visual prognosis in aniridia. About one third of the cases are sporadic, and two thirds are familial, with autosomal dominant inheritance and high penetrance. Nearly one third of sporadic AN patients develop Wilms tumor in association with

Internet zone

13. Cerca in questa sezione informazioni sulla funzione, sulla localizzazione e sull'espressione di Pax6. Quale organo colpiscono le malattie associate a mutazioni in questo gene? (risposta: gli occhi) Fai scorrere la pagina fino alla sezione Cross-references.

as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

### Cross-references

EMBL	M77844; AAA59963.1; -. [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] [ <a href="#">CoDingSequence</a> ]
	M77844; AAA59962.1; -. [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] [ <a href="#">CoDingSequence</a> ]
	M93650; AAA36416.1; -. [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] [ <a href="#">CoDingSequence</a> ]
	BC011953; AAH11953.1; -. [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ] [ <a href="#">CoDingSequence</a> ]
	Z83307; -. NOT_ANNOTATED_CDS. [ <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> ]
PIR	A56674; A56674.
PDB	6PAX; 13-JUL-99. [ <a href="#">ExPASy</a> / <a href="#">RCSB</a> / <a href="#">EBI</a> ]
TRANSFAC	T01122; -. [ <a href="#">TRANSFAC</a> ]
	T01814; -. [ <a href="#">TRANSFAC</a> ]
Genew	HGNC:8620; PAX6.
CleanEx	HGNC:8620; PAX6.
GeneCards	<a href="#">PAX6</a> .
GeneLynx	<a href="#">PAX6</a> ; Homo sapiens.
GenAtlas	<a href="#">PAX6</a> .
MIM	607108 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	106210 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	109120 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	120200 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	120430 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	129750 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	136520 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	148190 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
	165550 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].
604229 [ <a href="#">NCBI</a> / <a href="#">EBI</a> ].	
GO	<a href="#">GO:0003700</a> ; Molecular function: transcription factor activity ( <i>traceable author statement</i> ).
	<a href="#">GO:0007417</a> ; Biological process: central nervous system development ( <i>traceable author statement</i> ).
	<a href="#">GO:0007456</a> ; Biological process: eye morphogenesis (sensu Drosophila) ( <i>traceable author statement</i> ).

In questa sezione sono contenuti i link a banche dati che contengono informazioni di altro tipo riferite a Pax6. E' possibile anche ritrovare le stesse pagine eseguendo una ricerca nelle banche dati specifiche utilizzando i numeri d'accesso segnalati in questa pagina.

14. Prendi nota di tutti i numeri d'accesso della banca dati PDB, OMIM (segnalati con MIM) e SMART (che si trova piu' in basso in questa pagina). Fai scorrere la pagina fino a raggiungere la sezione Sequence information.

**Sequence information**

Length: **422** AA    Molecular weight: **46683** Da    CRC64: **C33CDD2C1B13C397** [This is a checksum on the sequence]

10	20	30	40	50	60
MQNSHSGVNO	LGGVFVNGRP	LPDSTRQKIV	ELAHSGARPC	DISRILQVSN	GCVSKILGRY
70	80	90	100	110	120
YETGSIRPRA	IGGSKPRVAT	PEVVSKIAQY	KRECPSIFAW	EIRDRLLESEG	VCTNDNIPSV
130	140	150	160	170	180
SSINRVLRLN	ASEKQQMGAD	GMYDKLRMLN	GQTGSWGTRP	GWYPGTSVPG	QPTQDGCQQQ
190	200	210	220	230	240
EGGENTNSI	SSNGEDSDEA	QMRQLQKRKL	QRNRTSFTQE	QIEALEKEFE	RTHYPDV FAR
250	260	270	280	290	300
ERLAAKIDLP	EARIQVWFSN	RRAKWRREEK	LRNQRRQASN	TPSHIPISSS	FSTSVYQPIP
310	320	330	340	350	360
QPTTPVSSFT	SGSMLGRDTR	ALTNTYSALP	PMPSTMANN	LPMQPPVPSQ	TSSYSCLMPT
370	380	390	400	410	420
SPSVNGRSYD	TYTPPHMQTH	MNSQPMGTSG	TTSTGLISPG	VSVVPQVPGS	EPDMSQYWRP

LQ

P26367 in [FASTA format](#)

Questa sezione riporta la sequenza della proteina Pax6 umana, puoi dedurre la lunghezza e il peso molecolare.

A questo punto le informazioni che hai annotato sul foglietto ti dicono già molte cose sulla proteina di cui fino a poco fa conoscevi solo la sequenza in zebrafish! Ma con i numeri di accesso di cui hai preso nota possiamo sapere molto di più visitando altre banche dati.

15. Apri l'home page dell'NCBI (vedi glossario) all'indirizzo <http://www.ncbi.nlm.nih.gov/> e fai click sulla parola TaxBrowser in bianco sulla striscia blu sotto il logo.

Entrez Taxonomy

Address: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy>

NCBI Entrez Taxonomy

Search  for

Limits Preview/Index History Clipboard Details

## The NCBI Entrez Taxonomy Homepage

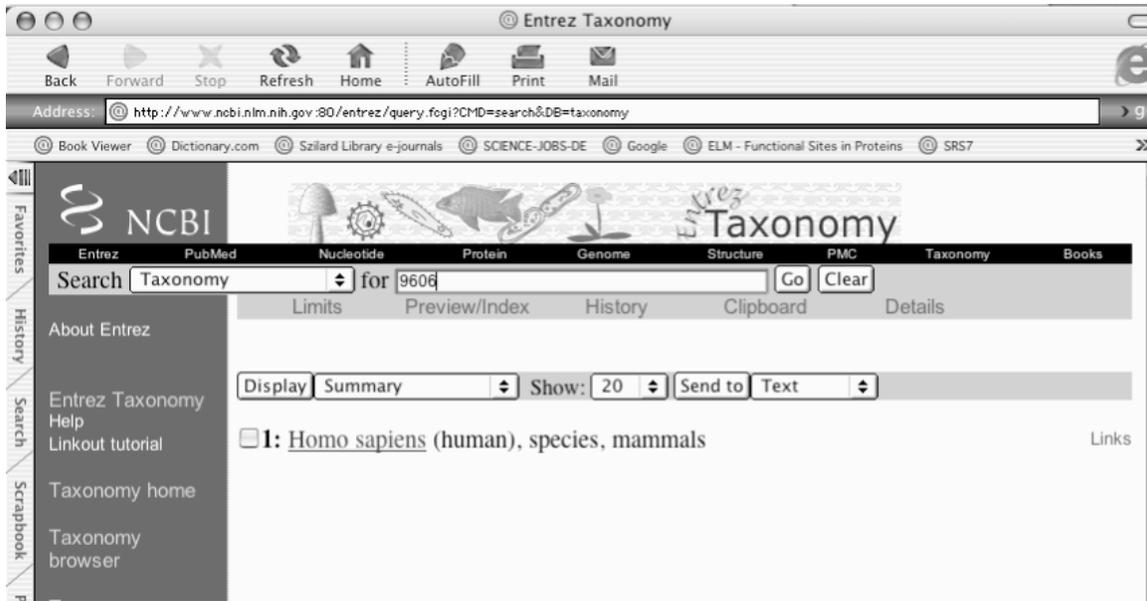
The NCBI taxonomy database contains the names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence. Click on the [tree](#) if you want to browse the taxonomic structure or retrieve sequence data for a particular group of organisms.

These are direct links to some of the organisms commonly used in molecular research projects:

<input type="checkbox"/> <a href="#">Arabidopsis thaliana</a>	<input type="checkbox"/> <a href="#">Hepatitis C virus</a>	<input type="checkbox"/> <a href="#">Pneumocystis carinii</a>
<input type="checkbox"/> <a href="#">Bos taurus</a>	<input type="checkbox"/> <a href="#">Homo sapiens</a>	<input type="checkbox"/> <a href="#">Rattus norvegicus</a>
<input type="checkbox"/> <a href="#">Caenorhabditis elegans</a>	<input type="checkbox"/> <a href="#">Magnaporthe grisea</a>	<input type="checkbox"/> <a href="#">Saccharomyces cerevisiae</a>
<input type="checkbox"/> <a href="#">Chlamydomonas reinhardtii</a>	<input type="checkbox"/> <a href="#">Mus musculus</a>	<input type="checkbox"/> <a href="#">Schizosaccharomyces pombe</a>
<input type="checkbox"/> <a href="#">Danio rerio (zebrafish)</a>	<input type="checkbox"/> <a href="#">Mycoplasma pneumoniae</a>	<input type="checkbox"/> <a href="#">Takifugu rubripes</a>
<input type="checkbox"/> <a href="#">Dictyostelium discoideum</a>	<input type="checkbox"/> <a href="#">Neurospora crassa</a>	<input type="checkbox"/> <a href="#">Xenopus laevis</a>
<input type="checkbox"/> <a href="#">Drosophila melanogaster</a>	<input type="checkbox"/> <a href="#">Oryza sativa</a>	<input type="checkbox"/> <a href="#">Zea mays</a>
<input type="checkbox"/> <a href="#">Escherichia coli</a>	<input type="checkbox"/> <a href="#">Plasmodium falciparum</a>	

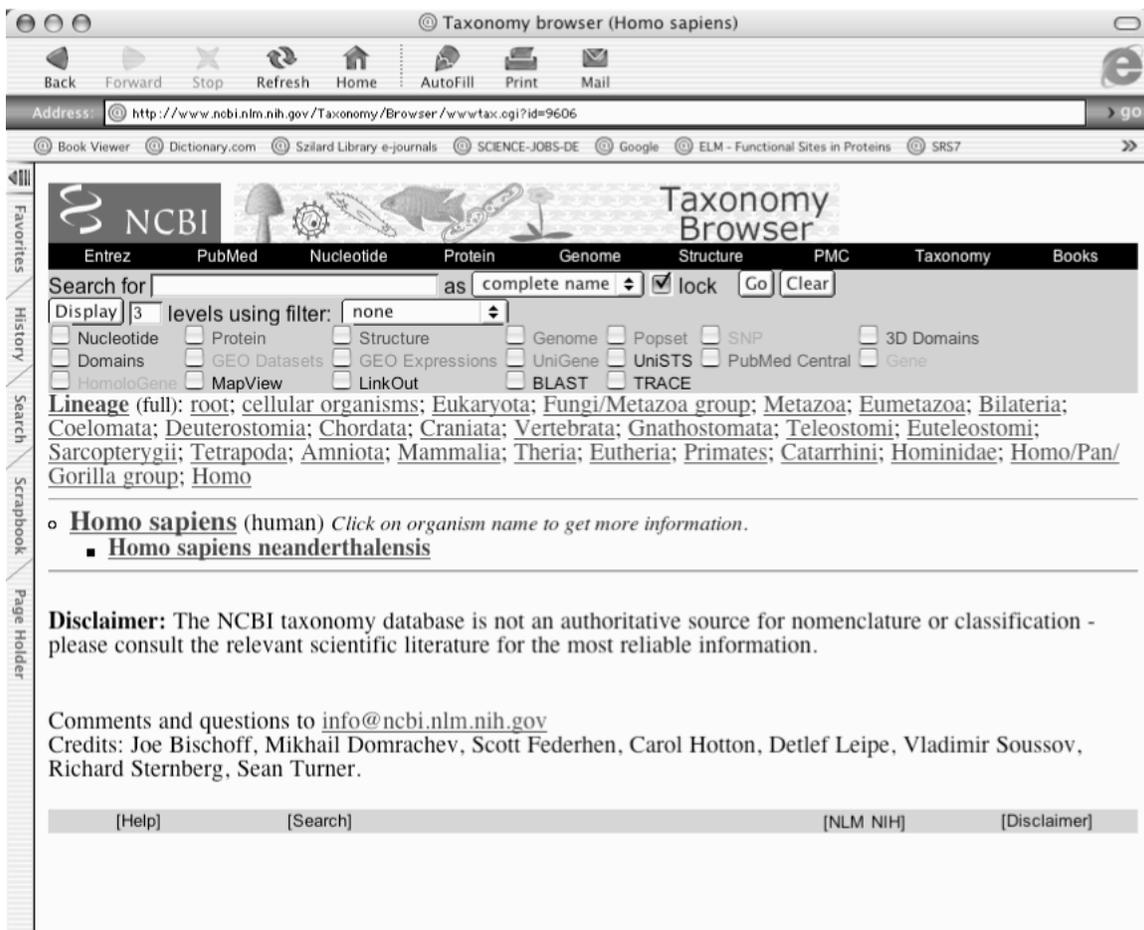
Comments and questions to [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

16. Nella casella per la ricerca digita il numero TaxID di cui hai preso nota (9606) e fai click sul tasto Go.



Come puoi vedere il numero di accesso corrisponde all'organismo *Homo sapiens*.

17. Per saperne di piu' fai click sulla parola *Homo sapiens*.



Ci sono due specie sotto *Homo sapiens*, una delle quali estinta!

18. Fai click sulla prima voce.

**Homo sapiens**

*Taxonomy ID:* 9606  
*Genbank common name:* **human**  
*Rank:* species  
*Genetic code:* Translation table 1 (Standard)  
*Mitochondrial genetic code:* Translation table 2 (Vertebrate Mitochondrial)  
*Other names:*  
 common name: **man**

*Lineage( full )*  
 cellular organisms; [Eukaryota](#); [Fungi/Metazoa](#)  
 group; [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#);  
[Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#);  
[Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#);  
[Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Primates](#);  
[Catarrhini](#); [Hominidae](#); [Homo/Pan/Gorilla](#) group; [Homo](#)

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	<a href="#">8,574,462</a>	<a href="#">8,574,456</a>
Protein	<a href="#">221,532</a>	<a href="#">221,532</a>
Structure	<a href="#">5,225</a>	<a href="#">5,225</a>
Genome	<a href="#">25</a>	<a href="#">25</a>
Popset	<a href="#">8,063</a>	<a href="#">8,063</a>
SNP	<a href="#">9,854,145</a>	<a href="#">9,854,145</a>
3D Domains	<a href="#">19,242</a>	<a href="#">19,242</a>
Domains	<a href="#">25</a>	<a href="#">25</a>
GEO Datasets	<a href="#">170</a>	<a href="#">170</a>
GEO Expressions	<a href="#">2,642,108</a>	<a href="#">2,642,108</a>
UniGene	<a href="#">106,937</a>	<a href="#">106,937</a>
UniSTS	<a href="#">144,886</a>	<a href="#">144,886</a>
PubMed Central	<a href="#">1,493</a>	<a href="#">1,493</a>
Gene	<a href="#">33,122</a>	<a href="#">33,122</a>
HomoloGene	<a href="#">17,499</a>	<a href="#">17,499</a>
Taxonomy	<a href="#">2</a>	<a href="#">1</a>

**Genome Information**

See the [NCBI Genome homepage](#)  
 Go to [NCBI genomic BLAST page for Homo sapiens](#)

Genome view: 24 chromosomes

Names [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#)

See the [Mitochondrion Genome](#)

In questa pagina puoi trovare altre informazioni sulla classificazione dell'organismo che ti interessa, oltre al suo nome comune e il numero di cromosomi che lo caratterizzano. Facendo click sul numero di un cromosoma e' possibile accedere alla sua mappa.

19. Apri la home page di SMART all'indirizzo <http://smart.embl-heidelberg.de/>.

The screenshot shows the SMART web interface. At the top, there's a navigation bar with buttons like Back, Forward, Stop, Refresh, Home, AutoFill, Print, and Mail. The address bar shows the URL <http://smart.embl-heidelberg.de/>. Below the navigation bar, there's a sidebar with links for Favorites, History, Search, Scrapbook, and Page Holder. The main content area is divided into two columns: "Sequence analysis" and "Architecture analysis".

**Sequence analysis:** This section explains that users can use either the Swissprot/Sptrembl/Ensembl sequence identifier (ID) or the protein sequence itself to request the SMART service. It features a "Sequence ID or ACC" input field and a "Sequence" text area. Below these are "Sequence SMART" and "Reset" buttons. A note mentions that HMMER searches are the default, and users can also find "Outlier homologues and homologues of known structure", "PFAM domains", "signal peptides", and "internal repeats".

**Architecture analysis:** This section allows users to search for proteins with combinations of specific domains in different species or taxonomic ranges. It includes a "Domain selection" section with an input field and a "Domain selection" button, with examples like "TyrKc AND SH3 AND NOT SH2" and "UNIQUE SH2". There's also a "GO terms query" section with an input field and a "GO Query" button, with an example "membrane AND signal transduction". The "Taxonomic selection" section has a dropdown menu set to "All" and an input field with examples "Dictyostelium discoideum" and "Porifera", along with a "Reset" button.

At the top right of the main content area, there's a "Number of SMART HMMs" box showing "667" and a "QUEUE STATUS" box showing "empty". A vertical navigation menu on the right side lists "Simple", "Modular", "Architecture", "Research", and "Tool".

20. Digita nella casella search il codice per la nostra proteina di cui hai preso nota (pax6\_human) e fai click sul tasto Go.

SMART - Simple Modular Architecture Research ...

Address: [http://smart.embl-heidelberg.de/smart/show\\_motifs.pl](http://smart.embl-heidelberg.de/smart/show_motifs.pl)

References  
 Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864  
 Letunic et al. (2004) *Nucleic Acids Res* **32**, D142-D144

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

S  
M  
A  
R  
T

Simple  
 Modular  
 Architecture  
 Research  
 Tool

Domains within the query sequence swissprot|P26367|PAX6 HUMAN

1 100 200

PAX HOX

Mouse over domain / undefined region to see the limits; click on it to go to further annotation; right-click to save whole protein as PNG image  
 Transmembrane segments as predicted by the *TMHMM2* program (▬), coiled coil regions determined by the *Coils2* program (—) and Segments of low compositional complexity, determined by the *SEG* program (—)

**Additional information**  
[Display other IDs, orthology and alternative splicing data for this sequence.](#)  
**Domain architecture analysis**  
[Display all proteins with similar domain organisation.](#)  
[Display all proteins with similar domain composition.](#)

In questa pagina e' possibile trovare informazioni sui domini che caratterizzano la nostra proteina: un dominio PAX e un omeodominio. Il tipo di domini presenti spesso caratterizza anche la famiglia a cui una proteina appartiene. Per esempio Pax6 e' il sesto membro della famiglia Pax.

21. Apri l'home page della banca dati PDB all'indirizzo <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/> .

© PDBsum - A database of the known 3D structure...

Address: <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>

Book Viewer Dictionary.com Szilard Library e-journals SCIENCE-JOBS-DE Google ELM - Functional Sites in Proteins SRS7

**PDBsum: A database of the known 3D structures of proteins and nucleic acids**

Home Browse Contact us Help

27,665 entries Includes 673 superseded entries

1bpy  
**Browse**

- PDB codes
- Het Groups
- Ligands
- Enzymes
- PROSITE
- Species
- Highlights **NEW**

Enter PDB code (4 characters) 6PAX Find Reset

Search string  Search

*To search all TITLE, HEADER, COMPND and SOURCE records in the PDB (eg to find a given protein by name), type the search-string above and click on **Search**. For more information on searching, click here.*

Internet zone

22. Nella casella search digita il codice PDB di cui hai preso nota consultando SwissProt e fai click su Find.

© PDBsum entry: 6pax

Address: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl

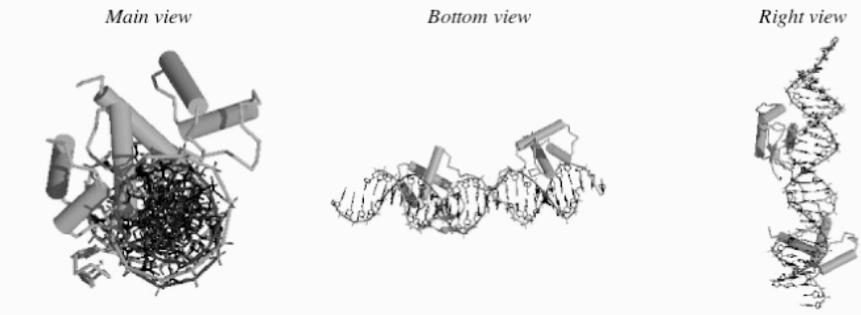
Book Viewer Dictionary.com Szilard Library e-journals SCIENCE-JOBS-DE Google ELM - Functional Sites in Proteins SRS7

EBI PDBsum

Top page Protein DNA/RNA Links

Gene regulation/DNA PDB id 6pax

Main view Bottom view Right view



Contents

- Description
  - Header details
  - Header records
- Protein chain
  - A 133 a.a. \*
- DNA/RNA
  - B
  - C
- Waters × 84

\* Residue conservation analysis

Go to PDB code: 6pax

PDB id: 6pax

Name: Gene regulation/DNA

Title: Crystal structure of the human pax-6 paired domain-DNA complex reveals a general model for pax protein-DNA interactions

Structure: Homeobox protein pax-6. Chain: a. Engineered: yes. Biological\_unit: monomer. 26 nucleotide DNA. Chain: b. Engineered: yes. Biological\_unit: monomer. 26 nucleotide DNA.

Source: Homo sapiens. Human. Gene: pax6. Expressed in: escherichia coll. Synthetic: yes. Synthetic: yes

SWISS-PROT: PAX6\_HUMAN [Pfam]

Resolution: 2.50Å

R-factor: 0.233

R-free: 0.256

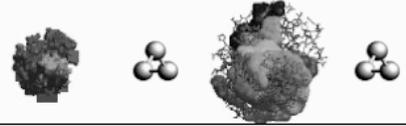
Authors: H.E.Xu,M.A.Rould,W.Xu,J.A.Epstein,R.L.Maas,C.O.Pabo

Date: 22-Apr-99

Links

- PDB
- MSD
- NDB
- MMDB
- Jena
- OCA
- CATH
- SCOP
- FSSP
- HSSP
- PQS
- ProSAT
- GRASS
- STING
- Procheck
- Whatcheck

Mol. surface and clefts:



WHATIF report on protein structural quality

Questa banca dati fornisce informazioni sulla struttura terziaria della proteina. Nella figura si vede il cristallo costituito dalla proteina associata al DNA.

23. Fai click sulla lettera A collegata alla parole Protein chain, sotto la voce Contents, sulla sinistra dello schermo.

Secondary structure: 6pax

Address: <http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=6pax&template=protein.html&l=1>

Book Viewer Dictionary.com Szilard Library e-journals SCIENCE-JOBS-DE Google ELM - Functional Sites in Proteins SRS7

EBI PDBsum

Top page Protein DNA/RNA Links

6pax

Protein chain A

Chain (133 residues)

SWISS-PROT code: PAX6\_HUMAN [Pfam]

**CATH** structural classification (2 domains):

Domain	Links	CATH no.	Class	Architecture
1	<a href="#">CATH</a> <a href="#">DHS</a>	1.10.10.10	= Mainly Alpha	Orthogonal Bundle
2	<a href="#">CATH</a> <a href="#">DHS</a>	"	= Mainly Alpha	Orthogonal Bundle

SHSGVNLGGVFNQRPDPSTRQRIVELAHSGARPCDISRI LQVSNQCYSKILGRYYAT  
5 10 15 20 25 30 35 40 45 50 55 60

GSIRPRAIGGSKPRVATPEVYSKIAQYKQECPSIFAWEIRDRLLEGVCTNDNIPSVSIS I  
65 70 75 80 85 90 95 100 105 110 115 120

NRVLRNLASEKQQ  
125 130

Residue interactions: with DNA

Analysis of sequence's residue conservation

Related protein sequences in the PDB

Protein chain A highlighted (click to view)

Motifs

- Secondary structure
  - Wiring diagram
  - Residue conservation
- ProMotif
  - 1 sheet
  - 2 strands
  - 6 helices
  - 9 beta turns
  - 1 gamma turn
  - 1 beta bulge
  - 1 beta hairpin
- PROSITE
  - PAIRED\_BOX

Go to PDB code:

Click to view

In questa pagina ci sono informazioni sulla struttura secondaria della proteina ed e' possibile vedere l'animazione della sua struttura tridimensionale facendo click sulla figura in alto a sinistra.

Ora che conosci le caratteristiche della proteina sei pronto a scoprire qualcosa di piu' sulla sua funzione e sulle malattie ad essa correlate.

24. Apri l'home page del sito NCBI <http://www.ncbi.nlm.nih.gov/> e fai click sulla parola OMIM in bianco sulla striscia blu sotto il logo.

OMIM – Online Mendelian Inheritance in Man

Address: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

NCBI  
 Online Mendelian Inheritance in Man  
 Johns Hopkins University

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search OMIM for pax6 human Go Clear

Limits Preview/Index History Clipboard Details

Entrez

- OMIM
- Search OMIM
- Search Gene Map
- Search Morbid Map

Help

- OMIM Help
- How to Link

FAQ

- Numbering System
- Symbols
- How to Print
- Citing OMIM
- Download

OMIM Facts

- Statistics
- Update Log
- Restrictions on Use

Allied Resources

- Genetic Alliance
- Databases
- HGMD
- Locus-Specific

**OMIM™ - Online Mendelian Inheritance in Man™**

Welcome to OMIM, Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information and references. It also contains copious links to MEDLINE and sequence records in the Entrez system, and links to additional related resources at NCBI and elsewhere.

You can do a search by entering one or more terms in the text box above. Advanced search options are accessible via the Limits, Preview/Index, History, and Clipboard options in the grey bar beneath the text box. The [OMIM help](#) document provides additional information and examples of basic and advanced searches.

25. Digita nella casella Search le parole Pax6 e human. L'utilizzo di due termini permette di restringere un po' la ricerca perche' il programma cercherà tutte le voci in banca dati che le contengano entrambe. Fai click su tasto Go.

The screenshot shows a web browser window with the address bar containing the URL: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Display&DB=omim>. The browser's address bar also shows several bookmarks: Book Viewer, Dictionary.com, Szilard Library e-journals, SCIENCE-JOBS-DE, Google, ELM - Functional Sites in Proteins, and SRS7.

The main content area of the browser displays the OMIM (Online Mendelian Inheritance in Man) website. The search bar contains the text "OMIM" and "for pax6 human". The search results are displayed in a table with two entries:

Item	Gene/Condition	Clinical Synopsis	GeneTests, Links
1	<b>*607108</b> <b>PAIRED BOX GENE 6; PAX6</b>	No clinical synopsis.	GeneTests, Links
2	<b>#106210</b> <b>ANIRIDIA, TYPE II; AN2</b>	<p><b>Clinical Synopsis</b></p> <p><b>INHERITANCE :</b></p> <p style="padding-left: 40px;">Autosomal dominant</p> <p><b>HEAD AND NECK :</b></p> <p style="padding-left: 40px;"><b>Eyes</b></p> <ul style="list-style-type: none"> <li>Aniridia</li> <li>Decreased vision</li> <li>Cataract</li> <li>Glaucoma</li> <li>Peter's anomaly (congenital anomaly of the anterior segment)</li> <li>Corneal clouding</li> </ul> <p><b>MOLECULAR BASIS :</b></p>	GeneTests, Links

The left sidebar of the website contains various navigation links such as "Entrez", "OMIM", "Search OMIM", "Search Gene Map", "Search Morbid Map", "Help", "FAQ", "OMIM Facts", "Allied Resources", and "Human Gene".

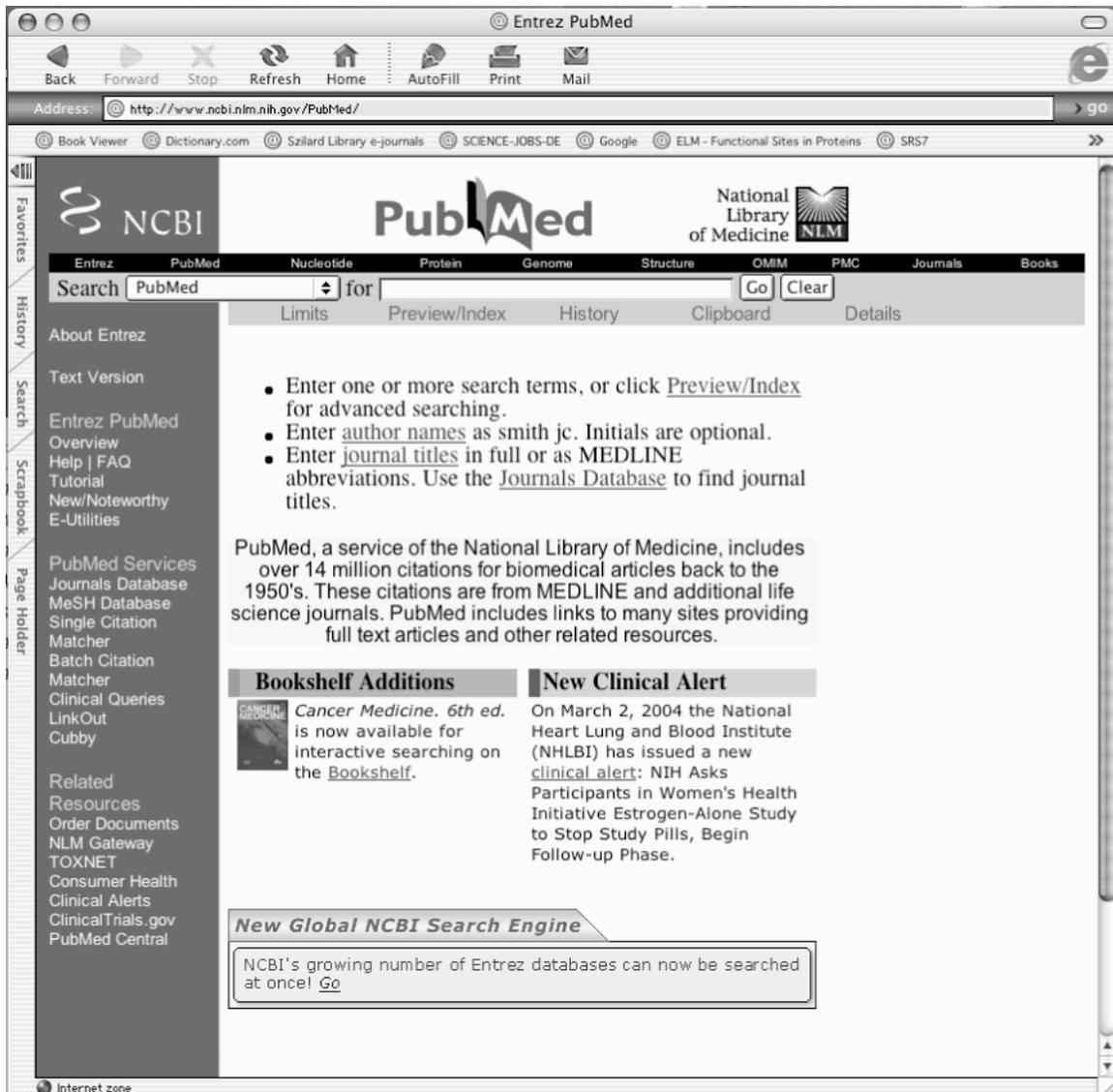
26. Fai click sulla prima voce

The screenshot shows the OMIM website interface. At the top, the browser title is "OMIM - PAIRED BOX GENE 6; PAX6". The address bar shows the URL: <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=607108>. The main content area displays the OMIM logo and the text "Online Mendelian Inheritance in Man" with the Johns Hopkins University logo. Below this is a search bar with "OMIM" entered and a "Go" button. There are also tabs for different data types: PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and OMIM. The main entry is for MIM \*607108, PAIRED BOX GENE 6; PAX6. The "DESCRIPTION" section states: "PAX6, a member of the paired box gene family, encodes a transcriptional regulator involved in oculogenesis and other developmental processes. For a discussion of paired box domain genes, see 167410." The "CLONING" section describes the discovery of a cDNA (D11S812E) in 1991. The "Gene map locus" is listed as 11p13.

Questa banca dati fornisce un riassunto dei dati principali legati ad una proteina, un gene o una malattia (vedi l'introduzione a OMIM). Confronta i numeri delle referenze che trovi citate con quelli di cui hai preso nota nella banca dati SwissProt. Alcune referenze che sono state trovate da questa ricerca non sono presenti nella tua lista. Si tratta di voci richiamate perché contengono le parole digitate ma che non si riferiscono specificamente alla proteina Pax6 (ad esempio la refrenza 167416).

Per ottenere informazioni più specifiche è possibile accedere direttamente alle pubblicazioni scritte dai ricercatori per comunicare il loro risultati. I riassunti (abstract) degli articoli sono raccolti nella banca dati PubMed (vedi introduzione a PubMed).

27. Apri l'home page del sito NCBI <http://www.ncbi.nlm.nih.gov/> e fai click sulla parola PubMed in bianco sulla striscia blu sotto il logo.



28. Nella casella Search digita Pax6 e fai click sul tasto Go.

Entrez PubMed

Search PubMed for PAX6

Display Summary Show: 20 Sort Send to Text

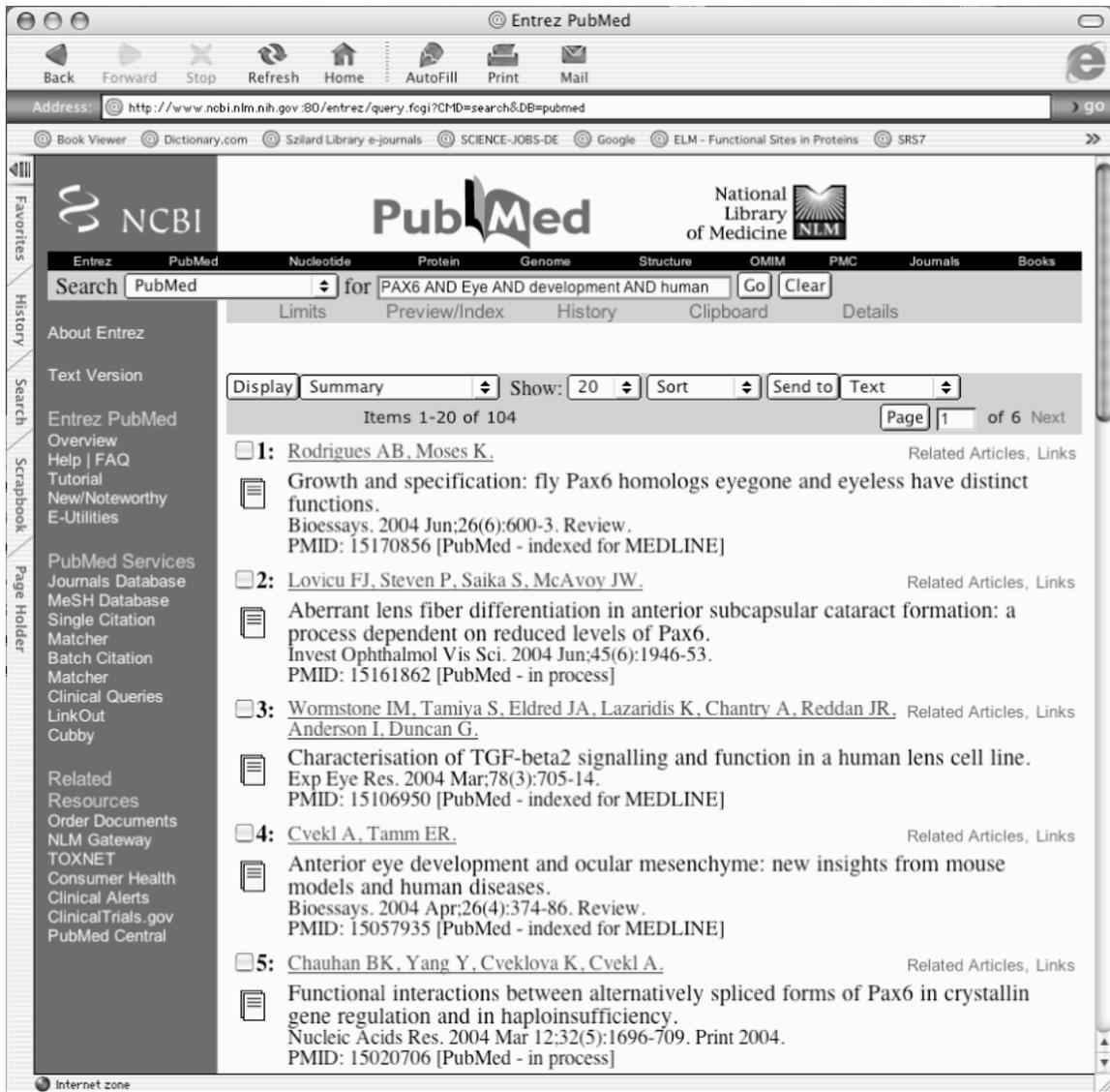
Items 1-20 of 746 Page 1 of 38 Next

- 1: Schuurmans C, Armant O, Nieto M, Stenman JM, Britz O, Klenin N, Brown C, Langevin LM, Seibt J, Tang H, Cunningham JM, Dyck R, Walsh C, Campbell K, Polleux F, Guillemot F. [Related Articles, Links](#)  
Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. *EMBO J.* 2004 Jul 1 [Epub ahead of print] PMID: 15229646 [PubMed - as supplied by publisher]
- 2: Schlosser G, Ahrens K. [Related Articles, Links](#)  
Molecular anatomy of placode development in *Xenopus laevis*. *Dev Biol.* 2004 Jul 15;271(2):439-66. PMID: 15223346 [PubMed - in process]
- 3: Bhattacharyya S, Bailey AP, Bronner-Fraser M, Streit A. [Related Articles, Links](#)  
Segregation of lens and olfactory precursors from a common territory: cell sorting and reciprocity of Dlx5 and Pax6 expression. *Dev Biol.* 2004 Jul 15;271(2):403-14. PMID: 15223343 [PubMed - in process]
- 4: Peng Y, Yang PH, Ng SS, Lum CT, Kung HF, Lin MC. [Related Articles, Links](#)  
Protection of *Xenopus laevis* Embryos Against Alcohol-induced Delayed Gut Maturation and Growth Retardation by Peroxiredoxin 5 and Catalase. *J Mol Biol.* 2004 Jul 16;340(4):819-27. PMID: 15223323 [PubMed - in process]
- 5: Ramon Martinez-Morales J, Rodrigo I, Bovolenta P. [Related Articles, Links](#)  
Eye development: a view from the retina pigmented epithelium. *Bioessays.* 2004 Jul;26(7):766-77. PMID: 15221858 [PubMed - in process]

Link: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\\_uids=15223343](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15223343)

Questa banca dati fornisce i riferimenti degli articoli pubblicati che contengano le parole con cui hai eseguito la ricerca (per saperne di più vedi l'introduzione a PubMed). I primi articoli in cima alla pagina sono, di solito, quelli più recenti. Facendo click sul simbolino giallo a sinistra è possibile accedere al riassunto dell'articolo e al testo completo dell'articolo, se l'editore lo mette a disposizione. Nella parte alta della pagina è scritto il numero totale di articoli presenti in banca dati che soddisfano la ricerca: 746. Per restringere la ricerca è possibile digitare più parole separate da "AND". Questo fa sì che il sistema cerchi solo gli articoli che contengono tutte le parole digitate.

29. Digita nella casella Search le parole "Pax6", "eye", "development", "human" separate da "AND".



In questo caso le voci sono "soltanto" 104 e sono piu' specifici per quanto riguarda lo sviluppo dell'occhio. Facendo click sull'icona gialla a sinistra del titolo potete accedere ai riassunti degli articoli.

A questo punto sai davvero quasi tutto quello che e' necessario per risolvere la domanda finale della caccia al tesoro!

## Una proteina nella rete: *La banca dati UniProt/Swiss-Prot*



UniProt/Swiss-Prot è una banca dati di sequenze proteiche che fornisce diversi tipi di informazioni: sequenze, riferimenti bibliografici, informazioni sull'organismo da cui deriva la proteina oltre a quelle che sono chiamate annotazioni (annotations). Le annotazioni comprendono la descrizione della funzione della proteina, le sue modificazioni post-trascrizionali (ad esempio carboidrati associati, fosforilazione, acetilazione), i suoi siti e i domini (come per esempio le regioni che legano il calcio, o i siti di legame per l'ATP, zinc fingers, homeobox), la sua struttura secondaria (alfa elica, beta foglietto), la struttura quaternaria (cioè se costituisce omodimeri, eterodimeri, ecc.), similarità con altre proteine, patologie associate a mutazioni nella sua sequenza, varianti...ecc.

Questo database è gestito in collaborazione dallo Swiss Institute for Bioinformatics (SIB) and the European Bioinformatics Institute (EBI).

L'homepage di UniProt/Swiss-Prot si trova all'indirizzo <http://www.expasy.org/sprot/>

The UniProt Knowledgebase consists of:

- **Swiss-Prot**: a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [[More details](#) / [References](#) / [Linking to Swiss-Prot](#) / [User manual](#) / [Recent changes](#) / [Commercial users](#) / [Disclaimer](#)].
- **TrEMBL**: a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at [SIB](#) and at [EBI](#).

**UniProt Release 2.0 consists of:**  
**Swiss-Prot Release 44.0 of 05-Jul-2004: 153871 entries** ([More statistics](#))  
**TrEMBL Release 27.0 of 05-Jul-2004: 1333917 entries** ([More statistics](#))

> **Swiss-Prot headlines**  
 New full release is available (44.0) ([Read more...](#))

**Access to Swiss-Prot and TrEMBL**

- [SRS](#) - Access to Swiss-Prot, TrEMBL and other databases using the Sequence Retrieval System
- [Full text search](#) in Swiss-Prot and TrEMBL
- [Advanced search in Swiss-Prot and TrEMBL](#) by description, gene name and organism (can be used to create html links to Swiss-Prot/TrEMBL queries)
- [Taxonomy browser \(NEWT\)](#)

Per eseguire una ricerca inserisci una o più termini nella casella "search for" e fai click sul pulsante d'azione "Go". Il sistema richiamerà tutte le informazioni che contengono i termini della ricerca. Per la ricerca puoi usare il nome o il simbolo di un

gene o di una proteina o i codici di accesso che hai ricavato da altre banche dati. Per esempio digita Pax6, seleziona Swiss-Prot come database e fai click sul pulsante Go.

Search in Swiss-Prot and TrEMBL for: pax6

Swiss-Prot Release 44.0 of 05-Jul-2004  
TrEMBL Release 27.0 of 05-Jul-2004

- Number of sequences found in [Swiss-Prot](#)<sub>(8)</sub> and [TrEMBL](#)<sub>(53)</sub>: **61**
- Note that the selected sequences can be saved to a file to be later retrieved; to do so, go to the [bottom](#) of this page.
- For more directed searches, you can use the Sequence Retrieval System [SRS](#).

**Search in Swiss-Prot: There are matches to 8 out of 153871 entries**

[PAX6\\_BRARE \(P26630\)](#)  
Paired box protein Pax[Zf-a] (Pax-6). {GENE: Name=pax6a; Synonyms=paxzf-a, pax[zf-a]} - Brachydanio rerio (Zebrafish) (Danio rerio)

[PAX6\\_CHICK \(P47237\)](#)  
Paired box protein Pax-6 (Fragment). {GENE: Name=PAX-6} - Gallus gallus (Chicken)

[PAX6\\_COTJA \(P47238\)](#)  
Paired box protein Pax-6 (Pax-QNR). {GENE: Name=PAX-6} - Coturnix coturnix japonica (Japanese quail)

[PAX6\\_DROME \(O18381\)](#)  
Paired box protein Pax-6 (Eyeless protein). {GENE: Name=ey; Synonyms=pax6; ORFNames=CG1464} - Drosophila melanogaster (Fruit fly)

[PAX6\\_HUMAN \(P26367\)](#)  
Paired box protein Pax-6 (Oculorhombin) (Aniridia, type II protein). {GENE: Name=PAX6; Synonyms=AN2} - Homo sapiens (Human)

[PAX6\\_MOUSE \(P32117\)](#)  
Paired box protein Pax-6 (Oculorhombin). {GENE: Name=Pax6; Synonyms=Pax-6, Sey} - Mus musculus (Mouse), Rattus norvegicus (Rat)

Il risultato della ricerca e' una lista di voci che riguardano Pax6 in diversi organismi. Nella prima riga c'e' il simbolo del gene in blu seguito dall'identificazione dell'organismo e da un codice che identifica questo gene. Nella seconda riga in nero c'e' il nome della proteina, in questo caso paired box protein Pax6, i suoi sinonimi e il nome dell'organismo di provenienza in inglese e in latino.

Fai click sulla voce che si riferisce a Pax6 umano.

**NiceProt View of Swiss-Prot: P26367**

Search  for

[Printer-friendly view](#) [Submit update](#) [Quick BlastP search](#)

[Entry info](#) [Name and origin](#) [References](#) [Comments](#) [Cross-references](#) [Keywords](#) [Features](#) [Sequence](#) [Tools](#)

*Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.*

**Entry information**

Entry name	PAX6_HUMAN
Primary accession number	P26367
Secondary accession number	Q99413
Entered in Swiss-Prot in	Release 23, August 1992
Sequence was last modified in	Release 38, July 1999
Annotations were last modified in	Release 44, July 2004

**Name and origin of the protein**

Protein name	Paired box protein Pax-6
Synonyms	Oculorhombin Aniridia, type II protein
Gene name	Name: PAX6 Synonyms: AN2
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**References**

[1] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=92103673;PubMed=1684738 [NCBI, ExPASy, EBI, Israel, Japan]  
Tay C C T, Hwang H, Miao H, Weil M M, Menachem B, Jordan T, van Heeringen V, Hestis

Link: <http://www.expasy.org/cgi-bin/aet-entries?RA=Hastie.N.D.>

Fai scorrere la pagina che hai ottenuto con la barra di scorrimento laterale: essa e' divisa in sezioni, dedicate ciascuna a un diverso tipo di informazioni e identificate dal titolo sulla striscia blu.

Sotto i primi due titoli, "Entry information" e "Name and origin of the protein", ritroviamo le informazioni sul nome della proteina, del gene e sulla posizione tassonomica dell'organismo di origine.

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

### References

- [1] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=92103673;PubMed=1684738 [NCBI, ExPASy, EBI, Israel, Japan]  
Ton C.C.T., Hirvonen H., Miwa H., Weil M.M., Monaghan P., Jordan T., van Heyningen V., Hastie N.D., Meijers-Heijboer H., Drechsler M., Royer-Pokora B., Collins F.S., Swaroop A., Strong L.C., Saunders G.F.;  
"Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region.";  
Cell 67:1059-1074(1991).
- [2] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=94258210;PubMed=1345175 [NCBI, ExPASy, EBI, Israel, Japan]  
Glaser T., Walton D.S., Maas R.L.;  
"Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene.";  
Nat. Genet. 2:232-239(1992).
- [3] SEQUENCE FROM NUCLEIC ACID (ISOFORM PAX6).  
**TISSUE=Lung**;  
DOI=10.1073/pnas.242603899;MEDLINE=22388257;PubMed=12477932 [NCBI, ExPASy, EBI, Israel, Japan]  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
Proc Natl Acad Sci U S A 99:16899-16903(2002)

Link: <http://www.expasy.org/cgi-bin/get-entries?RA=Dickson M.C.>

In questa sezione puoi trovare elencati i dati di alcuni articoli relativi alla proteina, Pax6.

**Comments**

- **FUNCTION:** Transcription factor with important functions in the development of the eye, nose, central nervous system and pancreas. Required for the differentiation of pancreatic islet alpha cells (*By similarity*). Competes with PAX4 in binding to a common element in the glucagon, insulin and somatostatin promoters (*By similarity*). The Pax6-5A isoform appears to function as a molecular switch that specifies target genes.
- **SUBCELLULAR LOCATION:** Nuclear.
- **ALTERNATIVE PRODUCTS:**
  - **Alternative splicing [3 named forms] Display all isoform sequences in Fasta format**  
*Comment:* Experimental confirmation may be lacking for some isoforms.

<b>Name</b>	<b>Pax6</b>
Isoform ID	P26367-1
This is the isoform sequence <a href="#">displayed in this entry</a> .	
<b>Name</b>	<b>Pax6-5A</b>
Isoform ID	P26367-2
Features which should be applied to build the isoform sequence: <a href="#">VSP_002366</a> .	
<b>Name</b>	<b>Pax6-5A,6*</b>
Isoform ID	P26367-3
The sequence of this isoform is not described.	

- **TISSUE SPECIFICITY:** Fetal eye, brain, spinal cord and olfactory epithelium. [Isoform Pax6-5A](#) is less abundant than the PAX6 shorter form.
- **DEVELOPMENTAL STAGE:** Expressed in the developing eye and brain.
- **DISEASE:** Defects in PAX6 are the cause of aniridia type II (AN2) [[MIM:106210](#)]. AN2 is a bilateral panocular disorder characterized by complete or partial absence of the iris, absence of the fovea and malformations of the lens and anterior chamber. Severe age-related corneal degeneration is a frequent complication which contributes to a poor visual prognosis in aniridia. About one third of the cases are sporadic, and two thirds are familial, with autosomal dominant inheritance and high penetrance. Nearly one third of sporadic AN patients develop Wilms tumor in association with

Nella sezione “Comments” si trovano informazioni sulla funzione della proteina e sulle eventuali malattie ad essa correlate. I codici in blu preceduti dalla lettera MIM costituiscono un link alle pagine web della banca dati di OMIM (vedi l’introduzione a OMIM).

as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

**Cross-references**

EMBL	M77844; AAA59963.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	M77844; AAA59962.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	M93650; AAA36416.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	BC011953; AAH11953.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	Z83307; -. NOT_ANNOTATED_CDS. [EMBL / GenBank / DDBJ]
PIR	A56674; A56674.
PDB	6PAX; 13-JUL-99.[ExpASY / RCSB / EBI]
TRANSFAC	T01122; -. T01814; -.
Genew	HGNC:8620; PAX6.
CleanEx	HGNC:8620; PAX6.
GeneCards	PAX6.
GeneLynx	PAX6; Homo sapiens.
GenAtlas	PAX6.
MIM	607108 [NCBI / EBI].
	106210 [NCBI / EBI].
	109120 [NCBI / EBI].
	120200 [NCBI / EBI].
	120430 [NCBI / EBI].
	129750 [NCBI / EBI].
	136520 [NCBI / EBI].
	148190 [NCBI / EBI].
165550 [NCBI / EBI].	
604229 [NCBI / EBI].	
GO	GO:0003700; Molecular function: transcription factor activity ( <i>traceable author statement</i> ).
	GO:0007417; Biological process: central nervous system development ( <i>traceable author statement</i> ).
	GO:0007456; Biological process: eye morphogenesis (sensu Drosophila) ( <i>traceable</i>

In questa sezione sono contenuti i link a banche dati che contengono informazioni di altro tipo riferite a Pax6, per esempio sulla struttura della proteina, sui suoi domini, sulle funzioni molecolari, ecc. E' possibile anche ritrovare le stesse pagine eseguendo una ricerca nelle banche dati specifiche utilizzando i numeri d'accesso segnalati in questa pagina. In particolare e' possibile accedere ai database OMIM, SMART e PDB (per avere altre informazioni, vedi la caccia al tesoro bioinformatica).

## Una proteina nella rete: La banca dati OMIM



OMIM, Online Mendelian Inheritance in Man, è una banca dati che contiene informazioni sui geni umani e sulle malattie genetiche realizzato e mantenuto dall'NCBI, the National Center for Biotechnology Information. Esso contiene la descrizione di geni e delle malattie ad essi associate, i quadri clinici e i riferimenti bibliografici, oltre a link a sequenze e ad altre risorse web. Si tratta della versione online del testo "Mendelian Inheritance in Man", ora alla sua dodicesima edizione, a cura di Victor A. McKusick e di un gruppo di colleghi della Johns Hopkins University e di altre istituzioni. La banca dati è aggiornata quotidianamente e a metà 2004 contiene oltre 15400 voci. Nella banca dati database sono riportate solo malattie che sono state associate ad uno o più geni.

La pagina di accesso ad OMIM si raggiunge tramite un link sulla homepage dell'NCBI: (<http://www.ncbi.nlm.nih.gov/>).

Fai click sulla parola OMIM in bianco sulla striscia blu in cima alla pagina

OMIM – Online Mendelian Inheritance in Man

Address: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

NCBI Online Mendelian Inheritance in Man Johns Hopkins University

Search OMIM for PAX6 Go Clear

Limits Preview/Index History Clipboard Details

- Enter one or more search terms.
- Use **Limits** to restrict your search by search field, chromosome, and other criteria.
- Use **Index** to browse terms found in OMIM records.
- Use **History** to retrieve records from previous searches, or to combine searches.

**OMIM™ - Online Mendelian Inheritance in Man™**

Welcome to OMIM, Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information and references. It also contains copious links to MEDLINE and sequence records in the Entrez system, and links to additional related resources at NCBI and elsewhere.

You can do a search by entering one or more terms in the text box above. Advanced search options are accessible via the Limits, Preview/Index, History, and Clipboard options in the grey bar beneath the text box. The [OMIM help](#) document provides additional information and examples of basic and advanced searches.

The links to the left provide further technical information, searching options, frequently asked questions ([FAQ](#)), and information on allied resources. To return to this page, click on the OMIM link in the black header bar or on the graphic at the top of any OMIM page.

Per eseguire una ricerca inserisci uno o più parole chiave nella casella “search for” a fai click sul pulsante “Go”. Il sistema cercherà tutte le voci in banca dati che contengano questi termini. Si possono usare il nome o il simbolo di un gene, il nome di una malattia, il numero di accesso ricavato da altre banche dati o qualunque altro termine che possa servire per richiamare le informazioni che ti interessano. Per esempio digita Pax6 e fai click sul pulsante “Go”.

The screenshot shows a web browser window with the address bar displaying `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=omim`. The page title is "OMIM - Online Mendelian Inheritance in Man". The search bar contains "OMIM" and "PAX6". The search results are displayed in a table format with the following entries:

OMIM ID	Gene Name	Gene Map Locus	Links
<a href="#">*607108</a>	PAIRED BOX GENE 6; PAX6	<a href="#">11p13</a>	GeneTests, Links
<a href="#">#106210</a>	ANIRIDIA, TYPE II; AN2	<a href="#">11p13</a>	GeneTests, Links
<a href="#">#136520</a>	FOVEAL HYPOPLASIA AND PRESENILE CATARACT SYNDROME FOVEAL HYPOPLASIA, ISOLATED, INCLUDED	<a href="#">11p13</a>	Links
<a href="#">#129750</a>	ECTOPIA PUPILLAE	<a href="#">11p13</a>	Links
<a href="#">#165550</a>	OPTIC NERVE HYPOPLASIA, BILATERAL OPTIC NERVE APLASIA, BILATERAL, INCLUDED	<a href="#">11p13</a>	Links
<a href="#">603807</a>	PETERS ANOMALY WITH CATARACT		GeneTests, Links

The page also shows a search bar with "OMIM" and "PAX6", a "Go" button, and a "Clear" button. The search results are displayed in a table format with columns for "Display", "Titles", "Show", "Send to", and "Text". The search results are numbered 1 through 6, and the first result is highlighted in blue. The search results are displayed in a table format with columns for "OMIM ID", "Gene Name", "Gene Map Locus", and "Links".

Questa pagina mostra una serie di voci che contengono i termini che hai digitato. Dal numero in alto puoi vedere che, in totale, le voci che soddisfano la tua ricerca sono 37. Nella prima riga c'è un codice in blu usato dal sistema di OMIM per identificare ciascuna voce della banca dati. Nella seconda riga, in nero, c'è il nome del gene, seguito dal suo simbolo. In questo caso il nome è paired box gene 6, che significa che si tratta del sesto membro della famiglia di proteine paired box, mentre il simbolo è Pax6. Nella terza riga in blu è indicata la localizzazione del gene. In questo caso 11p13 che significa che si trova sul cromosoma 11, nel braccio corto (p sta per petit, corto in francese), banda 13. In una stessa banda sono spesso presenti più geni, come si può vedere facendo click sulla localizzazione in blu. Fai click sulla prima voce che si riferisce specificamente a Pax6.

OMIM – PAIRED BOX GENE 6; PAX6

Address: <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=607108>

NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

Search OMIM for [ ] Go Clear

Display Detailed Show: 20 Send to Text

**\*607108** GeneTests, Links  
**PAIRED BOX GENE 6; PAX6**

Gene map locus [11p13](#)

**TEXT**

**DESCRIPTION**

PAX6, a member of the paired box gene family, encodes a transcriptional regulator involved in oculogenesis and other developmental processes. For a discussion of paired box domain genes, see [167410](#).

**CLONING**

Based on the map location of the aniridia type II ([106210](#)) locus, [Ton et al. \(1991\)](#) cloned a candidate cDNA (D11S812E) that was completely or partially deleted in 2 patients with aniridia. The smallest region of overlap between the 2 deletions, comprising less than 70 kb, encompassed the 3-prime coding region of the cDNA. This cDNA, which spanned over 50 kb of genomic DNA, detected a 2.7-kb message specifically within all tissues affected in aniridia. The predicted 422-amino acid polypeptide product possesses a paired domain, a homeodomain, and a serine/threonine-rich C-terminal domain, all structural motifs characteristic of certain transcription factors. All evidence pointed to D11S812E as being the AN2 gene.

The PAX6 gene encodes a transcriptional regulator that recognizes target genes through its paired-type DNA-binding domain. The paired domain is composed of 2 distinct DNA-binding

Dopo le informazioni relative al numero di identificazione, al nome e al simbolo del gene, si trova una parte di testo suddivisa in sezioni: Description, Cloning, Gene function, Molecular genetics, Animal Model, Allelic variants.

Ogni sezione e' costituita da un riassunto delle informazioni contenute negli articoli inerenti all'argomento, citati in blu. Facendo click sul nome dell'autore puoi accedere alla lista dei riferimenti bibliografici completi in fondo alla pagina. Il numero in blu alla fine di ogni referenza e' collegato tramite un link al riassunto del lavoro presente su PubMed.

Le lampadine gialle alla fine di ogni paragrafo permettono di acceder alla lista delle referenze citate in PubMed.

Per trovare informazioni su malattie associate a mutazioni in Pax6 leggi la sezione Molecular Genetics.

## Una proteina nella rete:

*La bancadati PubMed*



PubMed è una banca dati che permette di ottenere informazioni di tipo bibliografico nel campo della medicina e altre discipline di tipo biologico e naturalistico.

Gli editori della maggior parte delle riviste scientifiche internazionali inviano a PubMed una copia elettronica dei riferimenti bibliografici e, quasi sempre, anche del riassunto (abstract) di tutti gli articoli pubblicati in ogni numero della rivista. Inoltre, alcuni editori mettono a disposizione in rete anche il testo completo al quale PubMed permette l'accesso tramite un link. L'accesso al testo completo degli articoli può essere gratuito o a pagamento a seconda della casa editrice.

Il database contiene i riferimenti bibliografici a partire dagli anni '50 e viene aggiornato giornalmente. A metà del 2004 contiene circa 14 milioni di riferimenti bibliografici. Gli articoli provengono da riviste scientifiche di tutto il mondo ma la maggior parte delle voci sono in lingua inglese o hanno almeno il riassunto in inglese.

PubMed, come Omim, è un servizio che è stato creato e viene mantenuto a cura del National Center for Biotechnology Information (NCBI). È possibile accedere a PubMed dal sito NCBI (<http://www.ncbi.nlm.nih.gov/>).

The screenshot shows the NCBI homepage with the following elements:

- Navigation:** PubMed, Entrez, BLAST, OMIM, Books, TaxBrowser, Structure
- Search:** Search for  Go
- What does NCBI do?**

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)
- HIV-1 Protein Interaction Database**

HIV/AIDS researchers can now access a database of known interactions of HIV-1 proteins with proteins from human hosts. The database offers a concise summary of these interactions with links to PubMed, sequence data, and genes. [Read more...](#)
- Microbial Diversity**

Exploring "wild" microbial populations can reveal new species with novel metabolisms. A new study by Venter et al. describes a whole genome shotgun approach to environmental sampling of the Sargasso Sea. [Read more...](#)
- PubMed Central**
- Hot Spots:**
  - Clusters of orthologous groups
  - Coffee Break, Genes & Disease, NCBI Handbook
  - Electronic PCR
  - Entrez Home
  - Entrez Tools
  - Gene expression omnibus
  - Human genome resources
  - LocusLink
  - Malaria genetics & genomics
  - Map Viewer
  - dbMHC

Per aprire PubMed fai click sulla parola PubMed in bianco a sinistra sotto il logo NCBI.



Per richiamare informazioni su un argomento specifico digita una o più parole chiave nella casella “search for” e fai click sul pulsante “Go”. Il sistema cercherà le voci nel database che contengano queste parole. Come parole chiave puoi usare il nome o il simbolo di un gene o di una proteina, il nome di una malattia o altre parole in inglese che pensi possano essere utili per la tua ricerca. Per esempio digita il nome di una malattia: “aniridia”.

The screenshot shows a web browser window with the address bar displaying `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=pubmed`. The page header includes the NCBI logo, the PubMed logo, and the National Library of Medicine (NLM) logo. Below the header, there is a navigation menu with options like Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. The search bar contains the text 'Aniridia' and has 'Go' and 'Clear' buttons. Below the search bar, there are options for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The search results are displayed in a table with columns for 'Display', 'Summary', 'Show', 'Sort', and 'Send to'. The results are numbered 1 through 5, each with a yellow icon on the left. The authors' names are highlighted in blue. The titles of the articles and the journal names are also visible. The page number is 1 of 39.

Search results for 'Aniridia':

- 1:** [Mavrikakis I, Mavrikakis E, Svam PP, Bell J, Casey JH, Casswell AG, Brittain GP, Liu C.](#) Related Articles, Links  
Surgical management of iris defects with prosthetic iris devices. *Eye*. 2004 Jun 25 [Epub ahead of print] PMID: 15218523 [PubMed - as supplied by publisher]
- 2:** [Yalvac IS, Kocaoglan H, Eksioglu U, Demir N, Duman S.](#) Related Articles, Links  
Decompression retinopathy after Ahmed glaucoma valve implantation in a patient with congenital aniridia and pseudophakia. *J Cataract Refract Surg*. 2004 Jul;30(7):1582-5. PMID: 15210243 [PubMed - in process]
- 3:** [De Becker I, Walter M, Noel LP.](#) Related Articles, Links  
Phenotypic variations in patients with a 1630 A>T point mutation in the PAX6 gene. *Can J Ophthalmol*. 2004 Apr;39(3):272-8. PMID: 15180145 [PubMed - indexed for MEDLINE]
- 4:** [Laghmari M, Boutimzine N, Abdelouahed AK, Lezrek M, Bensouda N, Benharbit M, Ibrahimy W, Daoudi R, Benchrif Z, Mohcine Z.](#) Related Articles, Links  
[Bilateral congenital aniridia: 5 case reports] *J Fr Ophtalmol*. 2004 Apr;27(4):385-91. French. PMID: 15173647 [PubMed - indexed for MEDLINE]
- 5:** [Nagashima Y, Inayama Y, Kato Y, Sakai N, Kanno H, Aoki I, Yao M.](#) Related Articles, Links

Link: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Display&dopt=pubmed\\_pubmed&from\\_uid=15144395](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Display&dopt=pubmed_pubmed&from_uid=15144395)

La pagina web che viene aperta contiene i riferimenti bibliografici di tutti gli articoli presenti nel database che soddisfano la tua ricerca. Il nome degli autori appare in blu seguito dal titolo dell'articolo, il nome della rivista e la data di pubblicazione. Facendo click sull'icona gialla a sinistra puoi accedere al riassunto che ti interessa.

The screenshot shows the Entrez PubMed interface. At the top, there are navigation buttons (Back, Forward, Stop, Refresh, Home, AutoFill, Print, Mail) and a search bar with the URL: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\\_uids=15180145](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15180145). The main header includes the NCBI logo, the PubMed logo, and the National Library of Medicine (NLM) logo. Below the header, there are tabs for different database types: Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. The search bar contains the text "Search PubMed" and "for". There are buttons for "Go" and "Clear". Below the search bar, there are options for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The search results are displayed in a table with columns for "Display", "Abstract", "Show", "Sort", and "Send to". The first result is a paper by De Becker I, Walter M, and Noel LP, titled "Phenotypic variations in patients with a 1630 A>T point mutation in the PAX6 gene." The abstract text is visible below the title.

**Phenotypic variations in patients with a 1630 A>T point mutation in the PAX6 gene.**

De Becker I, Walter M, Noel LP.

Department of Ophthalmology, Dalhousie University, Halifax, NS.

**BACKGROUND:** The extreme clinical case-to-case variability of aniridia, even within families, can cause difficulties in making the correct diagnosis, prognosis and treatment plan. We describe seven patients from two families demonstrating variable expression of this syndrome, all with the same single point mutation within the PAX6 gene. **METHODS:** Case presentations. The authors review the clinical ophthalmic findings of the aniridia-keratopathy syndrome from two families, one involving four generations, the other with two generations. Polymerase chain reaction amplification of all 14 exons of the PAX6 gene was performed for five patients. **RESULTS:** The iris findings varied from classic total absence to nearly normal iris appearance. Corneal changes were minimal in the younger patients but vision-impairing in the older. None of the patients had glaucoma. The five patients who underwent DNA analysis shared the same PAX6 point mutation defect (1630A>T). **INTERPRETATION:** Bilateral corneal changes progressing from mild opacification at the limbus to vascularized central keratopathy, especially in the presence of nystagmus, are highly suggestive of a PAX6 mutation. Such patients may theoretically benefit from early limbal stem cell replacement therapy.

Il riassunto e' stato scritto dagli autori stessi per comunicare in breve il contenuto del loro articolo. In questa pagina, qualora sia disponibile, puoi trovare il link al testo completo.

Poiche' la maggior parte degli articoli sono molto specifici, esistono anche articoli che forniscono una rassegna sulle principali scoperte relative ad un argomento. Questo particolare tipo di articolo e' detto "review". Essi raccolgono le informazioni piu' aggiornate selezionando le piu' significative dagli articoli gia' pubblicati dagli stessi o da altri autori su un determinato argomento. Per limitare la tua ricerca a questo tipo di articoli torna alla pagina con la casella per la ricerca e fai click sulla parola in blu "Limits" sotto la casella per la ricerca. Apparira' la pagina seguente:

The screenshot shows the Entrez PubMed website interface. At the top, there is a navigation bar with links for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. Below this is a search bar with the text "Search PubMed for" and buttons for "Go" and "Clear". To the right of the search bar are links for "Limits", "Preview/Index", "History", "Clipboard", and "Details".

On the left side, there is a vertical menu with categories: Favorites, History, Search, Scrapbook, and Page Holder. Under "Search", there are links for "About Entrez", "Text Version", "Entrez PubMed Overview", "Help | FAQ", "Tutorial", "New/Noteworthy", and "E-Utilities". Under "Scrapbook", there are links for "PubMed Services", "Journals Database", "MeSH Database", "Single Citation Matcher", "Batch Citation Matcher", "Clinical Queries", "LinkOut", and "Cubby". Under "Page Holder", there is a "Related" link.

The main content area contains a list of search tips:

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- Search limits may exclude in process and publisher supplied citations.

Below the tips is a section titled "Limited to:" with several filters:

- All Fields (pull-down menu)
- only items with abstracts
- Review (pull-down menu)
- Languages (pull-down menu)
- Subsets (pull-down menu)
- Ages (pull-down menu)
- Human or Animal (pull-down menu)
- Gender (pull-down menu)
- Entrez Date (pull-down menu)
- Publication Date (pull-down menu) From [ ] [ ] [ ] [ ] To [ ] [ ] [ ] [ ]

Below the filters, there is a note: "Use the format YYYY/MM/DD; month and day are optional."

Seleziona "Review" dal menu' a tendina per il tipo di pubblicazione (Publication type). E' possibile anche selezionare altri limiti dai menu' disponibili per condizionare la tua ricerca. Digita di nuovo la parola "aniridia" e fai click sul pulsante "Go".

The screenshot shows the Entrez PubMed interface. The search bar contains 'PubMed' and the search term 'Aniridia'. The 'Limits' filter is checked, and the results are displayed as 'Review' articles. The search results list four items, each with a checkbox, a citation, and a 'Related Articles, Links' link.

**Search Results:**

- 1:** [Cvekl A, Tamm ER.](#) [Related Articles, Links](#)  
Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. *Bioessays*. 2004 Apr;26(4):374-86. Review. PMID: 15057935 [PubMed - indexed for MEDLINE]
- 2:** [Swanner JC, Walton DS, Chen TC.](#) [Related Articles, Links](#)  
Prevention of aniridic glaucoma with goniosurgery. *Int Ophthalmol Clin*. 2004 Winter;44(1):67-71. Review. No abstract available. PMID: 14704521 [PubMed - indexed for MEDLINE]
- 3:** [Krawczynski MR.](#) [Related Articles, Links](#)  
[Formation of congenital defects during ocular development. II. Genetic background of oculogenesis and developmental birth defects] *Klin Oczna*. 2003;105(3-4):207-10. Review. Polish. PMID: 14552188 [PubMed - indexed for MEDLINE]
- 4:** [Young TL.](#) [Related Articles, Links](#)  
Ophthalmic genetics/inherited eye disease. *Curr Opin Ophthalmol*. 2003 Oct;14(5):296-303. Review. PMID: 14502058 [PubMed - indexed for MEDLINE]

Gli articoli elencati in questa pagina sono tutte review sulla malattia aniridia.

## Una proteina nella rete:

### *Glossario*

**Abstract:** breve riassunto che illustra il contenuto di una pubblicazione.

**Accession number:** numero di serie assegnato a una sequenza nucleotidica o aminoacidica contenuti in una banca dati, in modo da poterla identificare in modo univoco.

**Bioinformatica:** disciplina cosituata dall'insieme degli strumenti informatici e delle loro applicazioni che sono utilizzati per ricerche di base e applicate nei diversi settori della biologia e della medicina.

**BLAST:** acronimo di Basic Local Alignment Search Tool. E' un programma per la ricerca di omologie locali di sequenza. BLAST può eseguire migliaia di confronti fra sequenze in pochi minuti perciò permette di confrontare in poco tempo una sequenza query con l'intero database per ricercare tutte le sequenze simili ad essa.

**Cross reference:** accession number ad altre banche dati che contengono informazioni relative al gene o alla proteina di interesse.

**Database:** banca dati, collezione di informazioni organizzate in modo che sia possibile richiamare diversi tipi di informazione (ad esempio PagineGialle on line).

**Dominio:** una parte della sequenza proteica con una struttura indipendente e con una funzione specifica

**E-value:** indica quanto è statisticamente probabile che l'allineamento tra due sequenze sia casuale. Più questo valore si avvicina allo zero più è probabile che le sequenze siano simili.

**Fattore di trascrizione:** una proteina che è coinvolta nella regolazione della trascrizione di un gene.

**Homepage:** pagina di accesso di un sito WEB. Dalla home page si possono trovare i link alle varie informazioni contenute nel sito.

**Link:** accesso diretto a diverse pagine di uno stesso sito Web o ad altri siti WEB (banche dati, referenze bibliografiche, ect) tramite un click del mouse su un'icona o una parola.

**Mutazione:** variazione nella sequenza nucleotidica di un gene che produce un fenotipo distinto.

**NCBI:** il National Center for Biotechnology Information, Bethesda, negli USA e' una fonte di informazioni di biologia molecolare. Questo centro crea e mantiene database pubblici,

svolge ricerche computazionali, sviluppa programmi per analizzare i dati genomici e si occupa della diffusione di informazione biomedica.

Omologia: il termine si riferisce a sequenze che condividono un progenitore comune. L'omologia e' un carattere qualitativo basato sulla misura oggettiva della similarita'.

Pax6: sesto elemento della famiglia di proteine Pax (=paired box), svolge la funzione di fattore di trascrizione.

Query: termine inglese che corrisponde a "interrogazione". Il dato (parole chiave o sequenze) che si immette nella apposita casella (indicata con "search for" o "find" o altro) costituisce la query.

Referenza: insieme dei dati che identifica una pubblicazione. Di solito comprende nome degli autori, titolo, nome della rivista o del libro e numeri di pagina.

Regioni a bassa complessita': sequenze proteiche che generalmente non sono considerate (mascherate) dalla ricerca perche' il loro allineamento avrebbe scarso significato biologico. Si tratta, per esempio di sequenze ricche di un dato amminoacido o di alcuni amminoacidi ripetuti piu' volte.

Sequenza amminoacidica: sequenza di amminoacidi che caratterizza una proteina.

Sequenza nucleotidica: sequenza di DNA o RNA composta appunto da nucleotidi.

Similarita': rappresenta un parametro quantitativo che esprime quanto una sequenza e' simile ad un'altra.