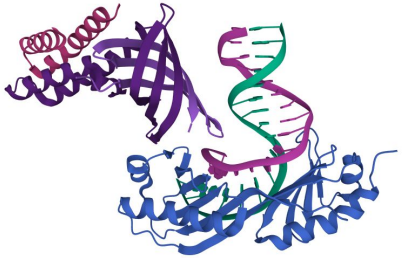
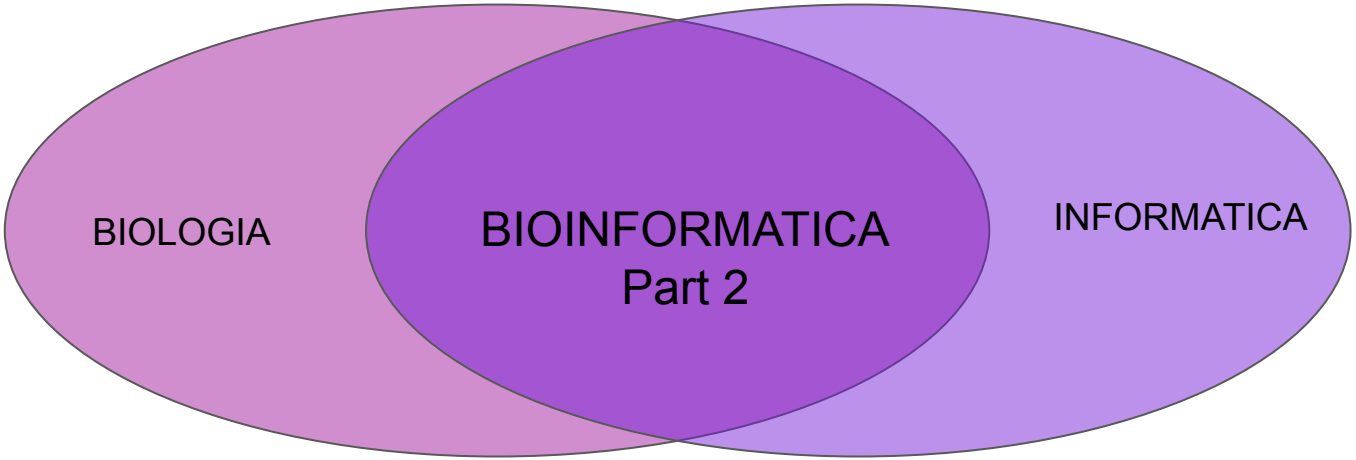
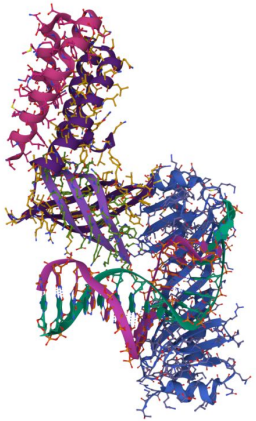
Annalaura Sabatucci





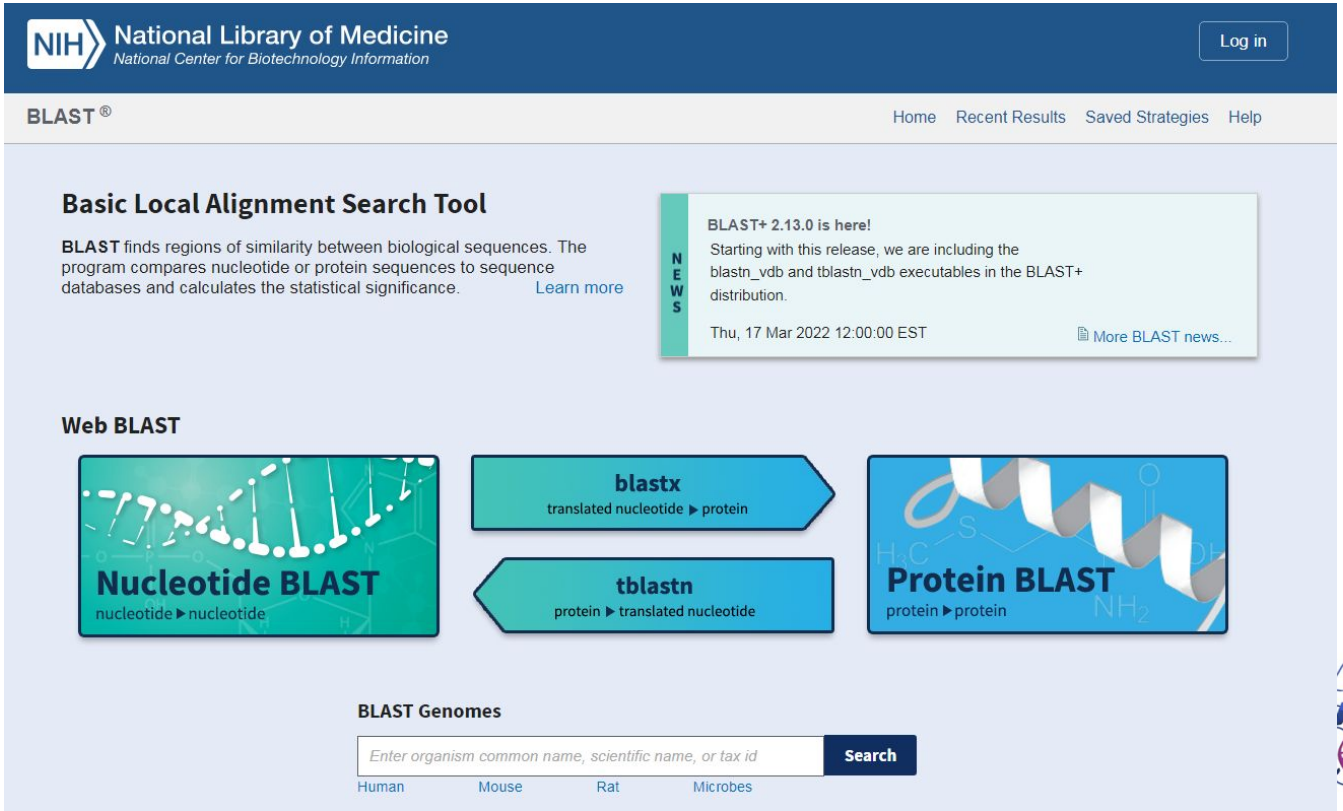
DEF: Approccio computazionale allo studio di sistemi biologici (a livello molecolare) => studio *in silico* (contrapposto allo studio *in vitro* o *in vivo*)

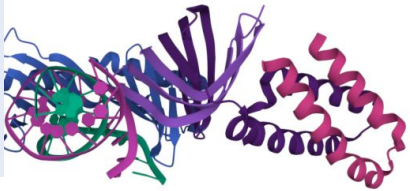
Alignment part 2: il BLAST



BLAST permette la RICERCA DI SEQUENZE SIMILI MEDIANTE ALLINEAMENTO SU INTERI DATABASES

https://blast.ncbi.nlm.nih.gov/Blast.cgi





Specialized searches

SmartBLAST

Find proteins highly similar to your query

Primer-BLAST

Design primers specific to your PCR template

Global Align

Compare two sequences across their entire span (Needleman-Wunsch)

CD-search

Find conserved domains in your sequence

IgBLAST

Search immunoglobulins and T cell receptor sequences

VecScreen

Search sequences for vector contamination

CDART

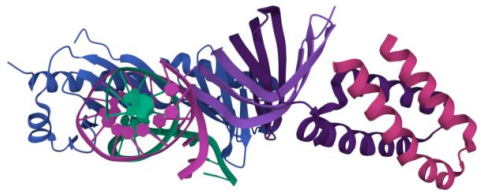
Find sequences with similar conserved domain architecture

Multiple Alignment

Align sequences using domain and protein constraints

MOLE-BLAST

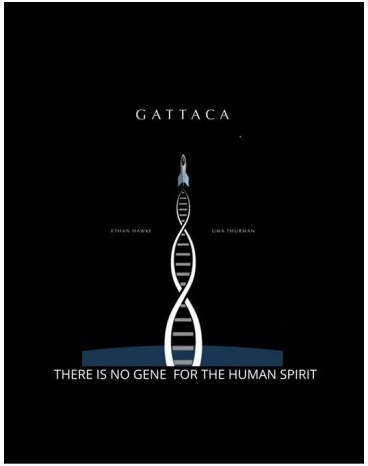
Establish taxonomy for uncultured or environmental sequences



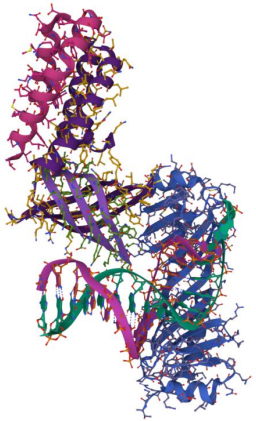
Sequence alignment software wikipedia page:

https://en.wikipedia.org/wiki/List_of_sequence_alignment_software

Have fun!



Studio dei genomi con EnsEMBL



GENE DELLA

Prodinorfina umana

Premessa

EnsEMBL stable Identifiers (IDs) : codici non ambigui e consistenti per identificare geni, trascritti, esoni ....

Stable ID Format

human: ENS[feature type prefix][a unique eleven digit number]

other species: ENS[species prefix][feature type prefix][a unique eleven digit number].

Feature Prefixes

E	exon
FM	Ensembl protein family
G	gene
GT	gene tree
P	protein
R	regulatory feature
T	transcript:

Species prefixes http://www.ensembl.org/info/genome/stable_ids/prefixes.html

Prefix	Species name
ENSGGO	Gorilla gorilla gorilla (Gorilla)
ENSMOC	Microtus ochrogaster (Prairie vole)
ENSPTR	Pan troglodytes (Chimpanzee)
ENSSCA	Serinus canaria (Common canary)
ENSOSI	Oryzias sinensis (Chinese medaka)
ENSCJA	Callithrix jacchus (White-tufted-ear marmoset)
MGP_CASTEiJ_	Mus musculus castaneus (Mouse CAST/EiJ)
ENSSSC	Sus scrofa (Pig - Bamei)
ENSCHI	Capra hircus (Goat)
ENSSRH	Sinocyclocheilus rhinocerous (Horned golden-line barbel)
ENSCMI	Callorhinchus milii (Elephant shark)

ES. ENST########## <=trascritto umano numero ##########

ENSMUSG########## <=gene, specie *Mus musculus* (topo) numero ##########

Prodynorphin:
human peptide hormone

UniProtKB - P01213 (PDYN_HUMAN)

Genetic variation and epigenetic modification of the prodynorphin gene in peripheral blood cells in alcoholism.

D'Addario C, Shchetynsky K, Pucci M, Cifani C, Gunnar A, Vukojević V, Padyukov L, Terenius L.

Prog Neuropsychopharmacol Biol Psychiatry. 2017 Jun 2;76:195-203. doi: 10.1016/j.pnpbp.2017.03.012. Epub 2017 Mar 21.

PMID: 28336495

The present study of human alcoholics aims to evaluate DNA methylation patterns in the prodynorphin gene (**PDYN**) promoter and to identify single nucleotide polymorphisms (SNPs) associated with alcohol dependence and with altered DNA methylation. ...Association with alcoholi ...

Regulation of gene transcription in bipolar disorders: Role of DNA methylation in the relationship between prodynorphin and brain derived neurotrophic factor.

D'Addario C, Palazzo MC, Benatti B, Grancini B, Pucci M, Di Francesco A, Camuri G, Galimberti D, Fenoglio C, Scarpini E, Altamura AC, Maccarrone M, Dell'Osso B.

Prog Neuropsychopharmacol Biol Psychiatry. 2018 Mar 2;82:314-321. doi: 10.1016/j.pnpbp.2017.08.011. Epub 2017 Aug 19.

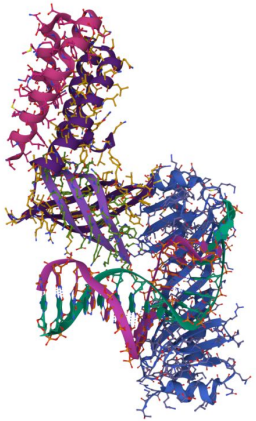
PMID: 28830794 Free PMC article.

Other target genes (i.e. catechol-O-methyltransferase (COMT), glutamate decarboxylase (GAD67), serotonin transporter (SERT) mRNA levels remained unaltered. Consistently, an increase in DNA methylation at **PDYN** gene promoter was observed in BD-II patients vs CT. After strati ...

IN EnsEMBL:

- Ricercare il gene umano
- annotare la posizione genica
- cercare numero di ortologhi e paraloghi
- andare al trascritto più rappresentativo
- visualizzare gli esoni
- customizzare il layout della pagina (es. variare il numero di bp visibili nelle regioni introniche)

Disegno di primers per pcr



TASKS:

- Vogliamo disegnare due primers per amplificare il gene della prodinorfina
- Vogliamo verificare che i primer amplifichino solo per il gene di interesse (BLAST)

Proprietà importanti di una buona coppia di primers

Ogni primers dovrà avere

- lunghezza basi compresa tra 18-24
- 40-60% G/C
- Distribuzione bilanciata di basi G/C e A/T
- T_m che permette un annealing tra 55-65° C
- NO strutture secondarie interne (hair-pins)

Le coppie di primers inoltre dovrebbero avere:

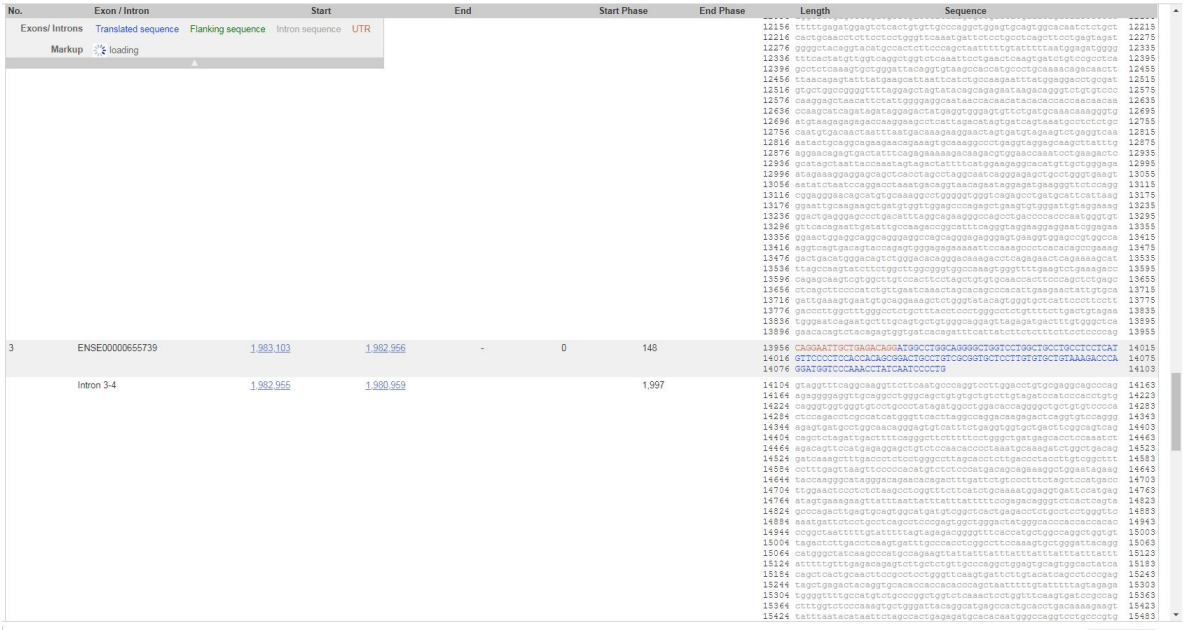
- T_m *simile* (*max* 2-3 ° C di differenza)
- **NO** complementarietà (> 2-3 bp) in particolare al 3'

SOFTWARE CONSIGLIATO
(ma anche qui ne esistono una grande varietà)

PRIMER3:

https://primer3.ut.ee/

Dal gene in ensEMBL selezioniamo il primo esone, a partire dal TSS (Transcription Start Site) : l'ATG che sodifica per la metionina iniziatrice



Primer3web version 4.1.0 - Pick primers from a DNA sequence.

disclaimer | code
cautions

Select the Task for primer selection generic

Template masking before primer design (available species)

Select species Example: Mus musculus | Nucleotides to mask in 5' direction 1
Primer failure rate cutoff < 0.1 | Nucleotides to mask in 3' direction 0

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a Mispriming Library (repeat library) NONE

☑ Pick left primer, or use left primer below | ☐ Pick hybridization probe (internal oligo), or use oligo below | ☑ Pick right primer, or use right primer below (5' to 3' on opposite strand)

Pick Primers | Download Settings | Reset Form

Sequence Id		A string to identify your output.
Targets		E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the source sequence with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.
Overlap Junction List		E.g. 27 requires one primer to overlap the junction between positions 27 and 28. Or mark the source sequence with -: e.g. ...ATCTAC-TGTCAT.. means that primers must overlap the junction between the C and T.
Excluded Regions		E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the source sequence with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.
Pair OK Region List		See manual for help.
Included Region		E.g. 20,400: only pick primers in the 400 base region starting at position 20. Or use { and } in the source sequence to mark the beginning and end of the included region: e.g. in ATC{TTC...TCT}AT the included region is TTC...TCT.
Start Codon Position		
Internal Oligo Excluded Region		

Force Left Primer Start -1000000 | Force Right Primer Start -1000000
Force Left Primer End -1000000 | Force Right Primer End -1000000

Sequence Quality

Min Sequence Quality 0 | Min End Sequence Quality 0 | Sequence Quality Range Min 0 | Sequence Quality Range Max 100

Pick Primers | Download Settings | Reset Form

General Primer Picking Conditions

Primer3 Output

```
PRIMER PICKING RESULTS FOR

Template masking not selected
No mispriming library specified
Using 1-based sequence positions
OLIGO            start  len      tm      gc%   any_th  3'_th hairpin seq
LEFT PRIMER         26   19   59.70   57.89     0.00   0.00    0.00 CTGCCTGCCTCCTCATGTT
RIGHT PRIMER       128   21   58.13   47.62     0.00   0.00    0.00 AGGGGATTGATAGGTTTGGGA
SEQUENCE SIZE: 129
INCLUDED REGION SIZE: 129

PRODUCT SIZE: 103, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00

    1 ATGGCCTGGCAGGGGCTGGTCCTGGCTGCCTGCCTCCTCATGTTCCCCTCCACCACAGCG
                              >>>>>>>>>>>>>>>>>>>

   61 GACTGCCTGTCGCGGTGCTCCTTGTGTGCTGTAAAGACCCAGGATGGTCCCAAACCTATC
                                                 <<<<<<<<<<<<<

  121 AATCCCCTG
      <<<<<<<<<

KEYS (in order of precedence):
>>>>>> left primer
<<<<<< right primer


Statistics
         con   too    in    in   not          no    tm    tm   high  high  high        high
         sid  many   tar  excl    ok   bad    GC   too   too any_th 3'_th hair-  poly   end
        ered    Ns   get   reg   reg   GC% clamp   low  high  compl compl   pin     X  stab    ok
Left     180     0     0     0     0   106     0     0    60      0     0     1     0     0    13
Right    180     0     0     0     0     0     0    64    20      0     0     0     0     0    96
Pair Stats:
considered 4102, unacceptable product size 4101, primer in pair overlaps a primer in a better pair 5264, ok 1
libprimer3 release 2.4.0


(primer3_results.cgi release 4.1.0)
```

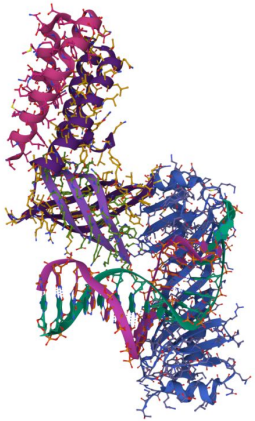
FINE I PARTE

- verifichiamo con BLAST i 2 primers

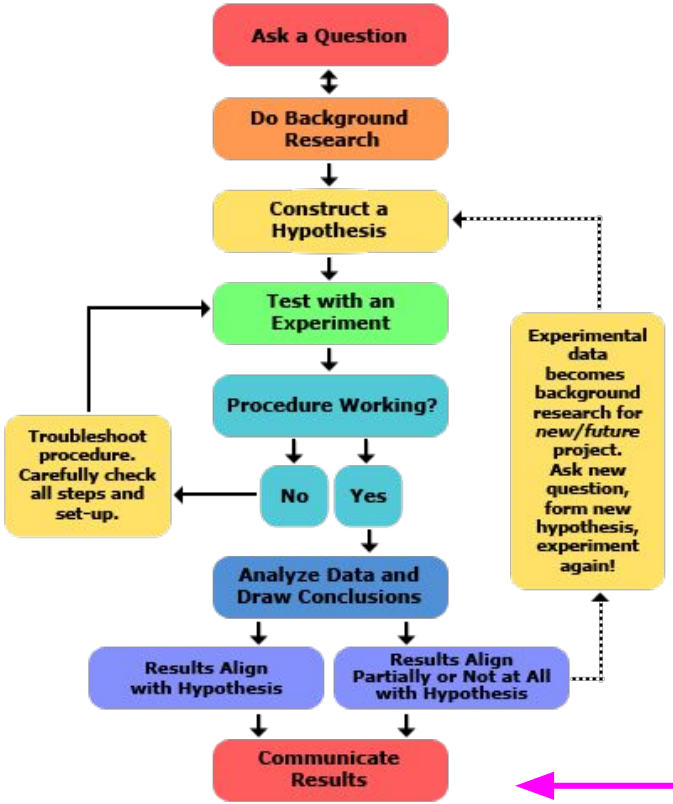
quale specie ci interessa?
Quale tipo di DNA? solo gli esoni?

FINE I PARTE

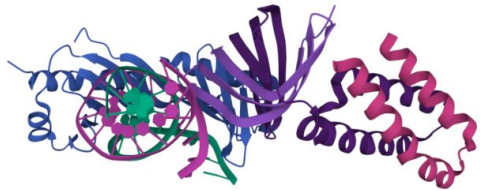
BIBLIOMETRIA



FLOWCHART DEL METODO SCIENTIFICO



La comunicazione dei risultati è la parte conclusiva e più importante di un lavoro.
Il mezzo di elezione è la pubblicazione su riviste scientifiche



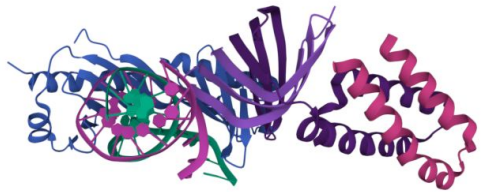
DATABASES DI BIBLIOGRAFIA SCIENTIFICA (riviste e libri)

PUBMED medicina e biologia

GOOGLE SCHOLAR

SCOPUS

WEB OF KNOWLEDGE



La **Bibliometria** fornisce parametri quantitativi sul 'valore' di una rivista scientifica o di un autore.

2 PARAMETRI BIBLIOMETRICI

RIVISTE:

Journal Impact Factor (IF)

E' una misura del numero di volte in cui gli articoli di una rivista vengono citati.

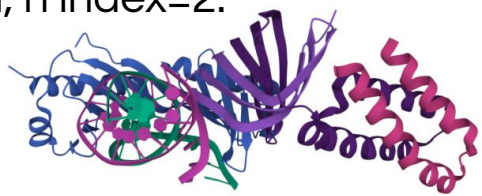
Maggiore IF, maggior prestigio della rivista

(Journal Citation Reports; SciMago Journal Rank)

AUTORI:

Author h-index

è una misura del numero di volte in cui il lavoro di un autore viene citato da altri.
Se un autore ha almeno 2 articoli citati almeno da 2 altri autori, h index=2.
Questo parametro viene riportato ad es. su SCOPUS.



ES. Cerchiamo su pubmed gli articoli pubblicati sulla prodinorfina dal prof D’Addario