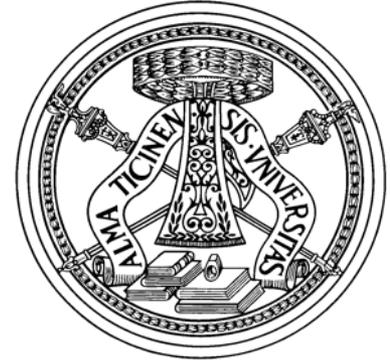
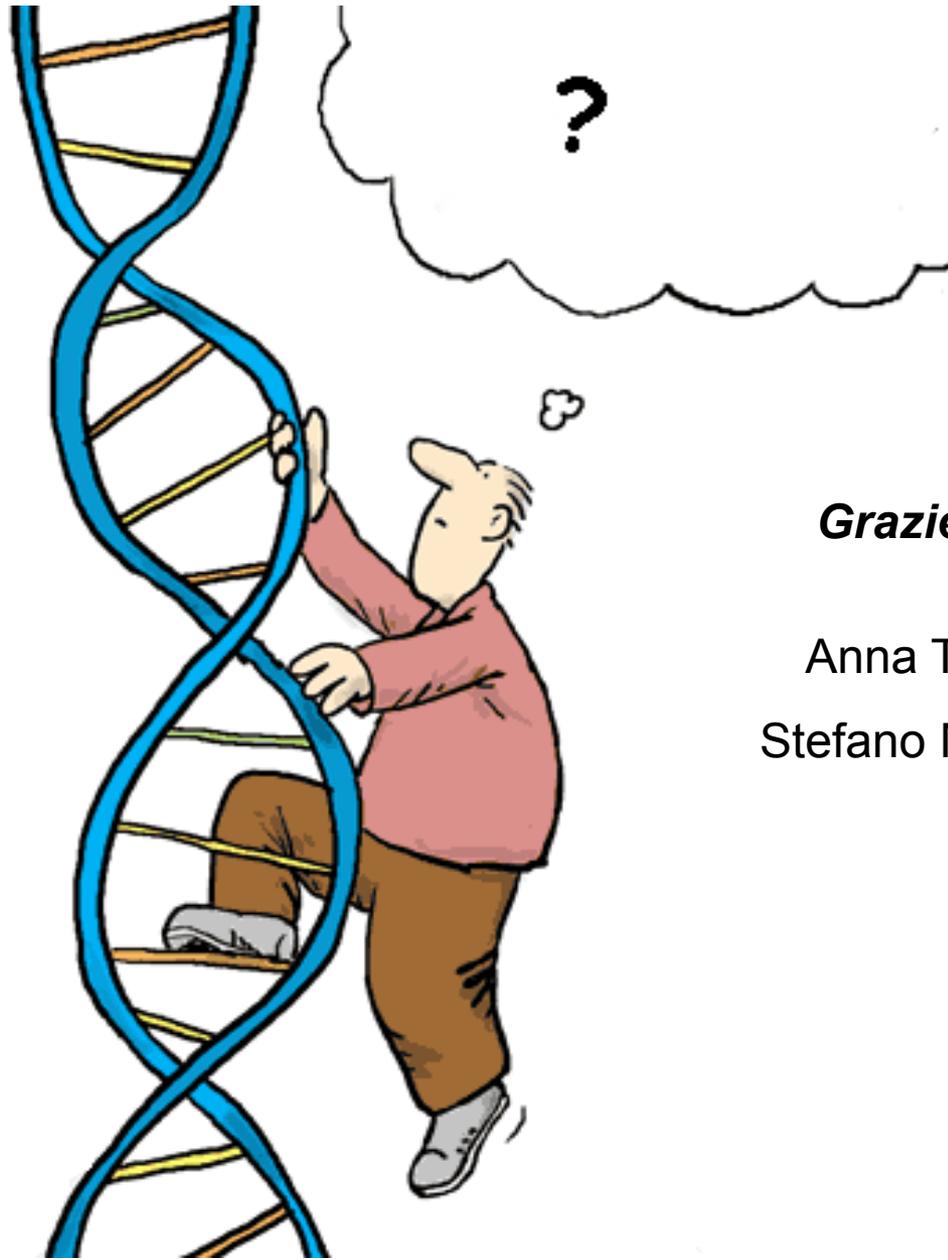


Giochi cooperativi e biomedicina

Giulia Cesari

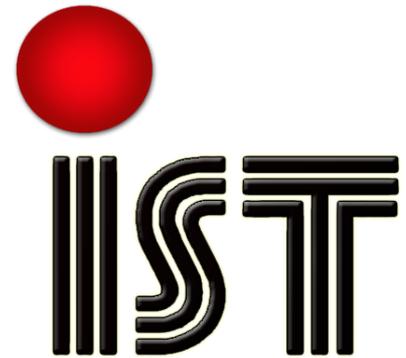
Politecnico di Milano e Université Paris Dauphine
giulia.cesari@polimi.it

Almo Collegio Borromeo, 2 maggio 2017



Grazie a:

Anna Torre
Stefano Moretti



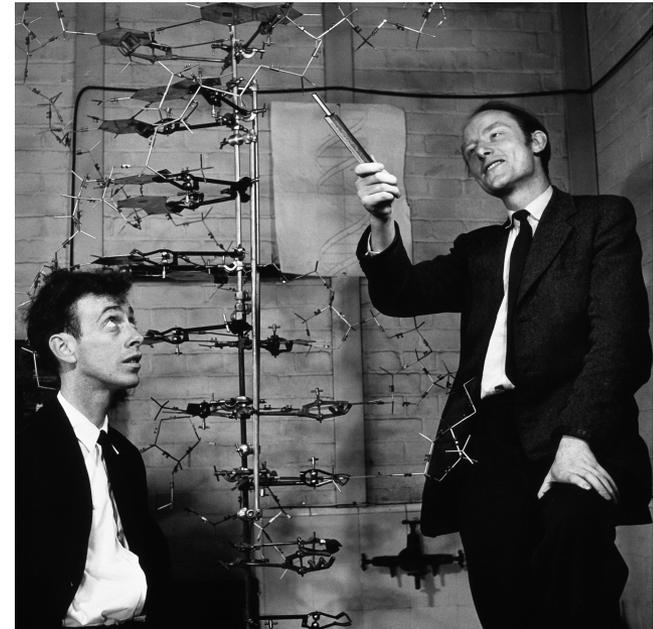
Neuroblastoma

- E' un tumore che ha origine dalle cellule del sistema nervoso autonomo
- Colpisce soprattutto neonati e bambini al di sotto dei 10 anni
- E' uno dei più diffusi tra i tumori dell'infanzia
- Nel 50% dei casi al momento della diagnosi è già metastatico e refrattario alla chemioterapia
- L'Associazione Italiana per la Lotta al Neuroblastoma nasce a Genova nel 1993.



Biologia in pillole

- Ogni organismo è costituito da *cellule* (nel corpo umano se ne stimano circa 6×10^{13})
- Le *proteine* costituiscono la struttura delle cellule e agiscono da enzimi nelle reazioni biochimiche
- I *geni* sono responsabili della *sintesi* delle proteine, il processo di produzione di una proteina sulla base dell'informazione codificata attraverso il DNA.



Francis Crick and James Watson
(February 28th, 1953)

Premio Nobel per la medicina, 1962

My Dear Michael,

Jim Watson and I have probably made a most important discovery. We have built a model for the structure of des-oxy-ribose-nucleic-acid (read it carefully) called D.N.A. for short. You may remember that the genes of the chromosomes — which carry the hereditary factors — are made up of protein and D.N.A.

Our structure is very beautiful. D.N.A. can be thought of roughly as a very long chain with flat bits sticking out. The flat bits are called the “bases”. [...]

Now we have two of these chains winding round each other — each one is a helix — and the chain, made up of sugar and phosphorus, is on the outside, and the bases are all on the inside. [...]

Now the exciting thing is that while these are 4 different bases, we find we can only put certain pairs of them together. The bases have names. They are Adenine, Guanine, Thymine & Cytosine. I will call them A, G, T and C. Now we find that the pairs we can make — which have one base from one chain joined to one base from another — are only A with T and G with C.

Now on one chain, as far as we can see, one can have the bases in any order, but if their order is fixed, then the order on the other chain is also fixed. [...]

It is like a code. If you are given one set of letters you can write down the others.

Now we believe that the D.N.A. is a code. That is, the order of the bases (the letters) makes one gene different from another gene (just as one page of print is different from another). You can now see how Nature makes copies of the genes. Because if the two chains unwind into two separate chains, and if each chain then makes another chain come together on it, then because A always goes with T, and G with C, we shall have two copies where we had one before. [...]

In other words, we think we have found the basic copying mechanism by which life comes from life. The beauty of our model is that the shape of it is such that only these pairs can go together, though they could pair up in other ways if they were floating about freely. You can understand that we are very excited. We have to have a letter off to Nature in a day or so.

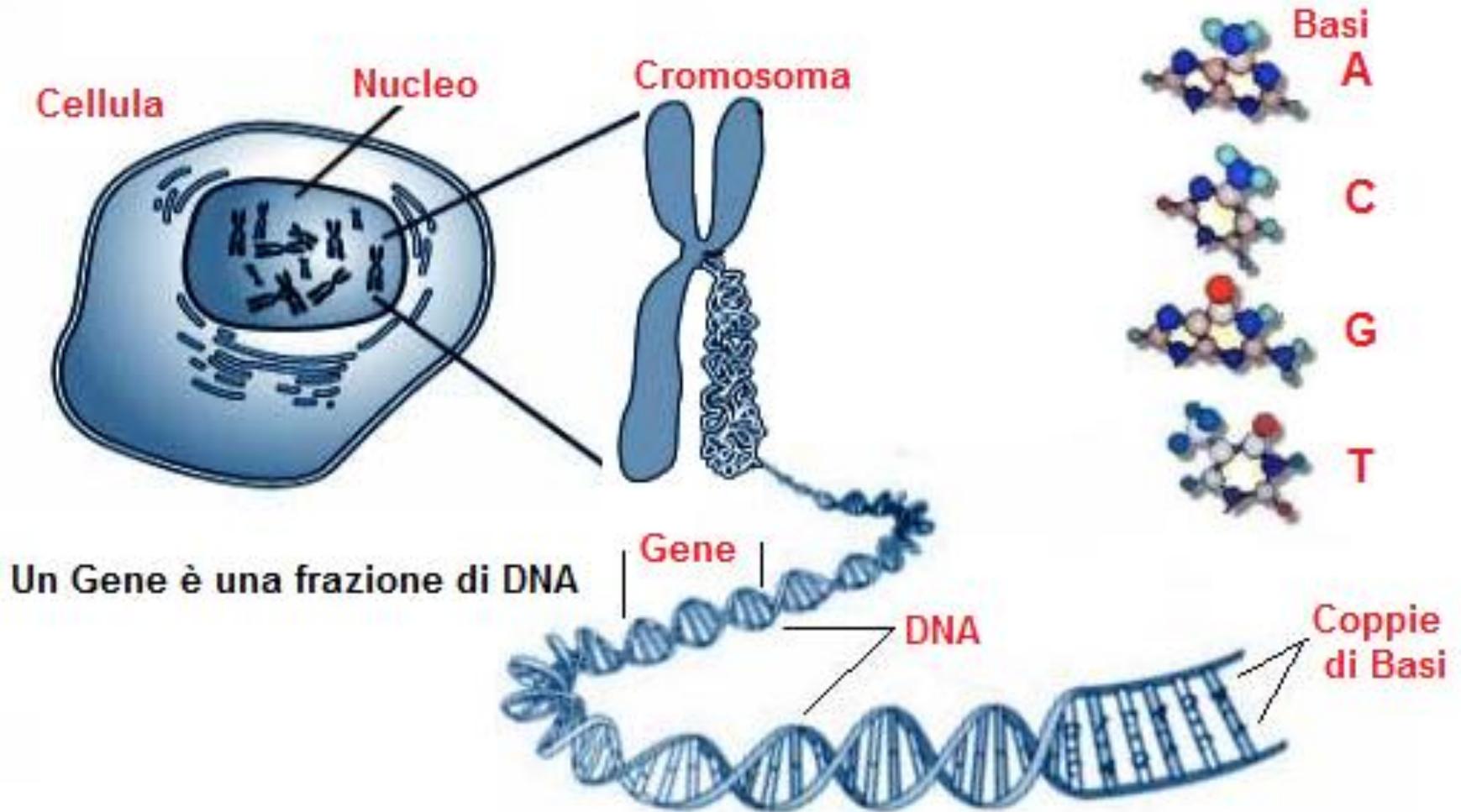
Read this carefully so that you understand it. When you come home we will show you the model.

Lots of love,

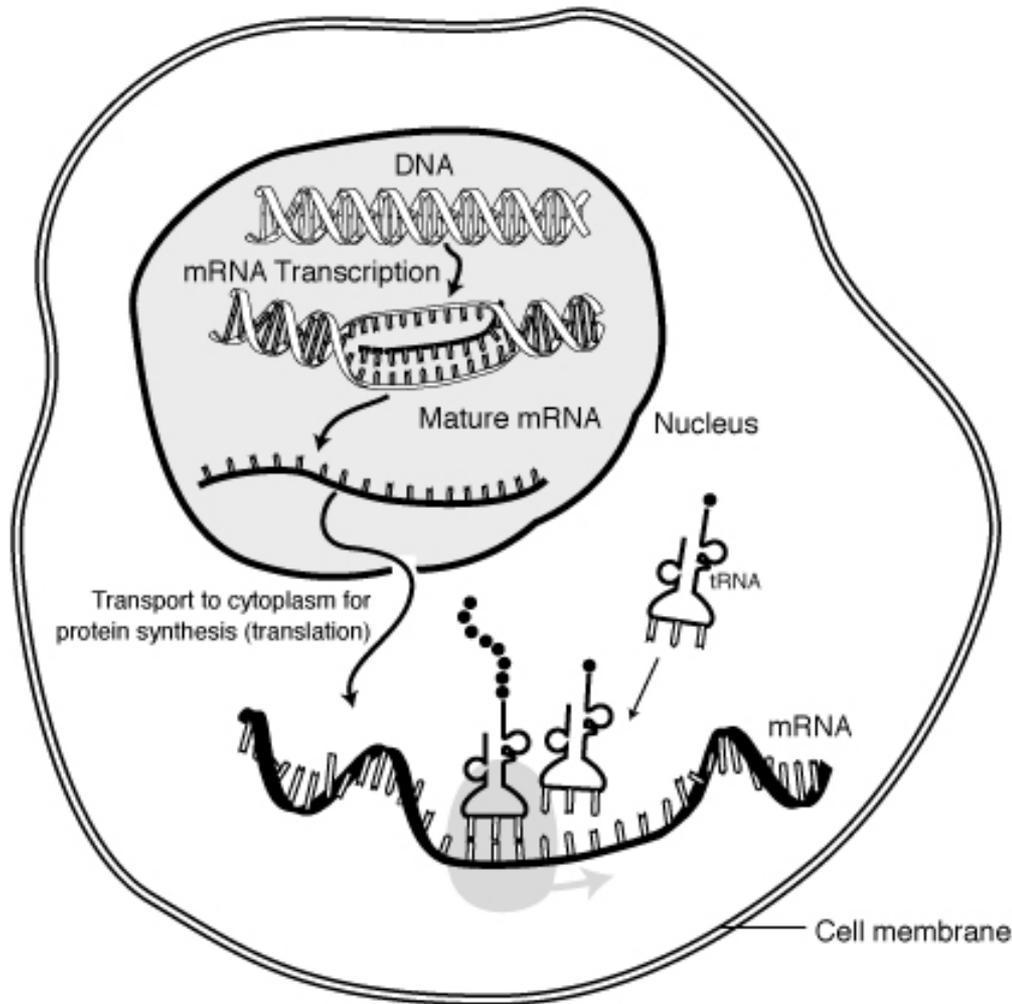
Daddy

March 19th, 1953, Francis Crick

I geni risiedono nella doppia elica del DNA, che contiene migliaia di geni.



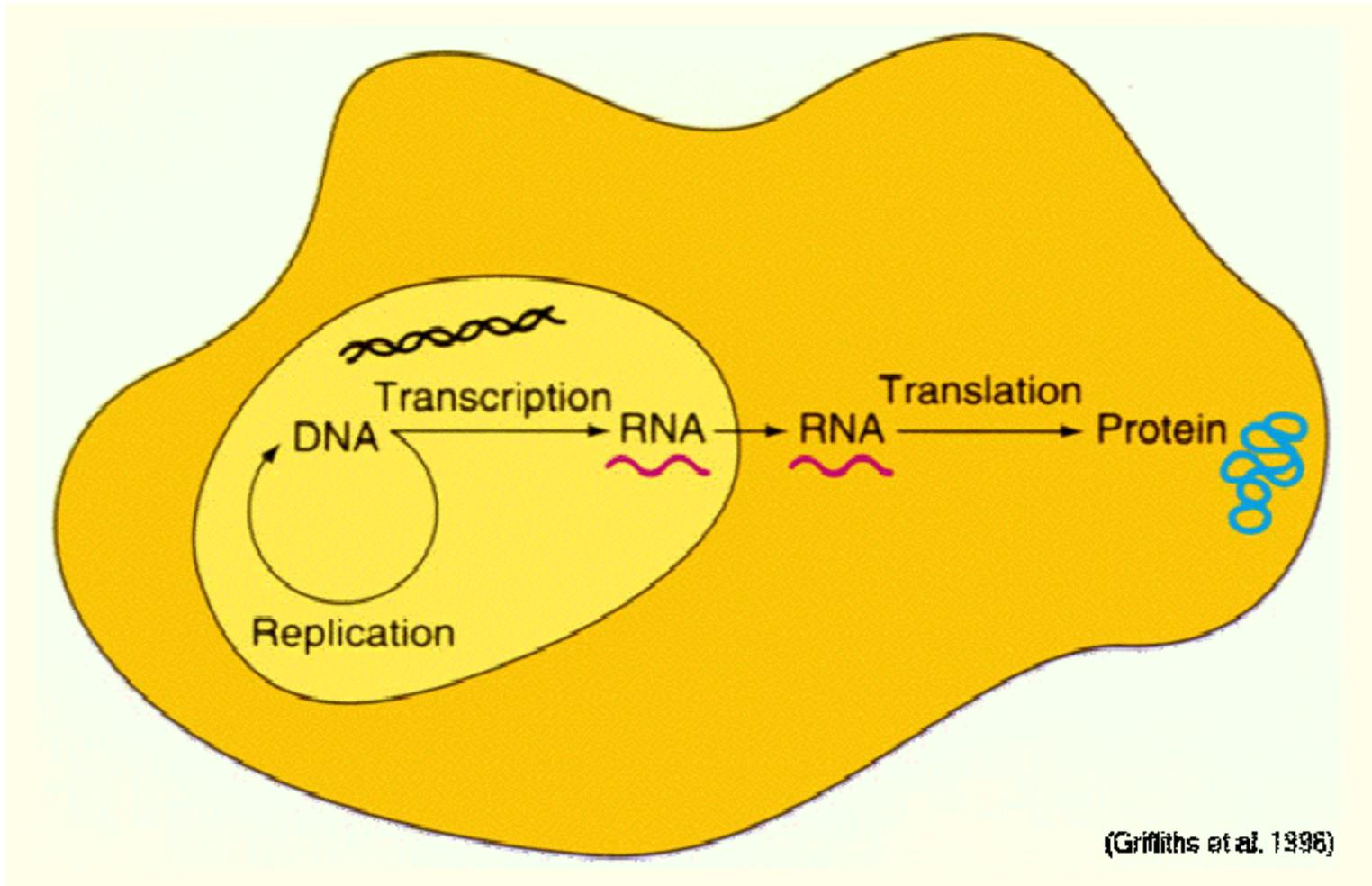
Dogma centrale della biologia molecolare



L'espressione di un gene si verifica quando l'informazione contenuta nel DNA viene:

- **trascritta** nelle molecole di acido ribonucleico messaggero (mRNA) e poi
- **tradotta** nelle proteine

I meccanismi di interazione tra geni, proteine e molecole all'interno di una cellula sono molto complessi ed è importante ricostruirli per comprendere la funzione dei geni all'interno di una condizione biologica d'interesse.



Misurare direttamente la quantità di proteine potrebbe essere meglio, ma al momento è difficile da realizzare.

Idea: misurare la quantità di **mRNA** per vedere quanti **geni** risultano essere **espressi** (“lavorano”) all’interno della cellula

Microarray

A cosa servono?

- Consentono la quantificazione simultanea dell'espressione di migliaia di geni

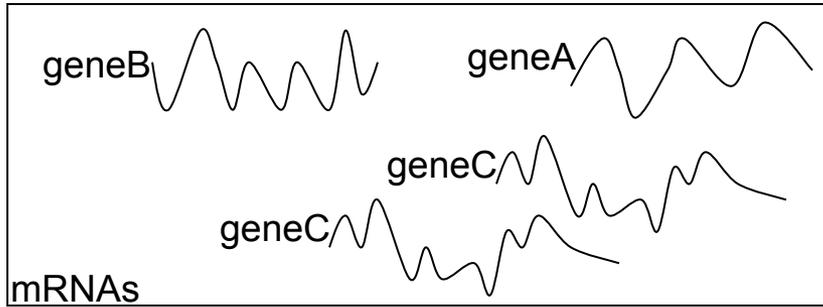
Cosa sono?

- Sono vetrini con moltissime sonde, ciascuna costituita da un segmento di DNA di un gene

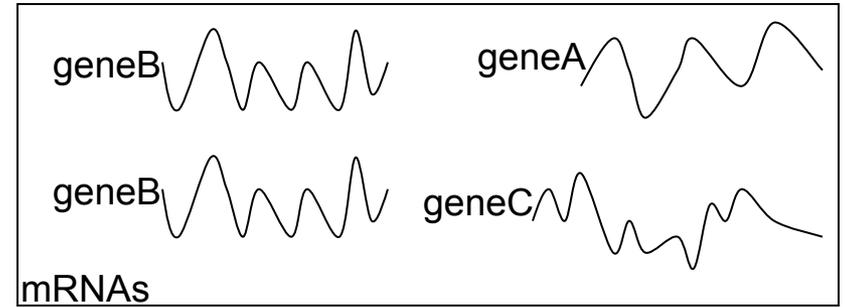
Come funzionano?

- Sfruttano la capacità delle molecole di mRNA di legarsi al filamento di DNA che lo ha generato, attraverso l'appaiamento di basi complementari (T-A, G-C)

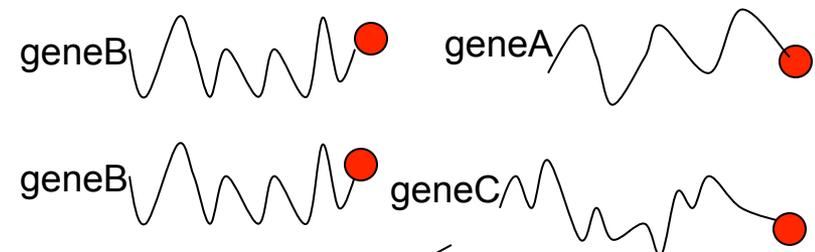
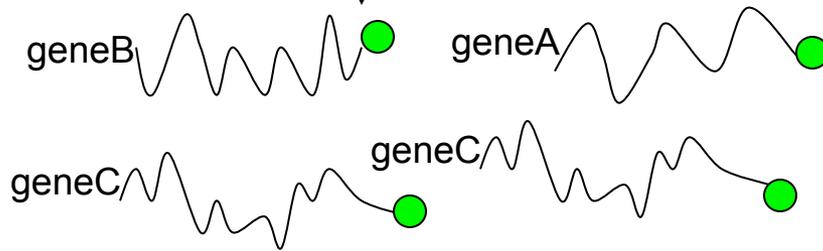
Cellula Normale



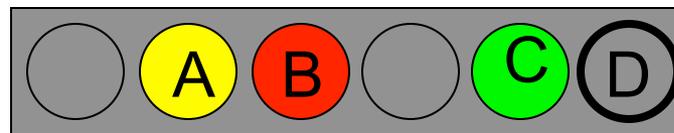
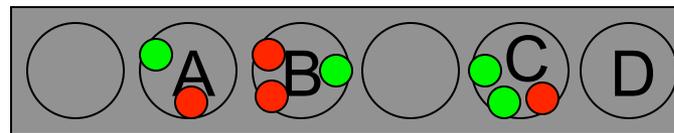
Cellula Tumorale



Reazione di marcatura fluorescente
mediante trascrizione inversa

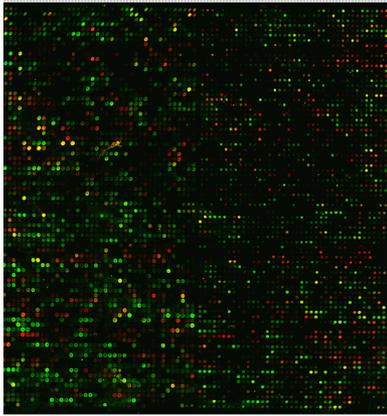


Ibridizzazione

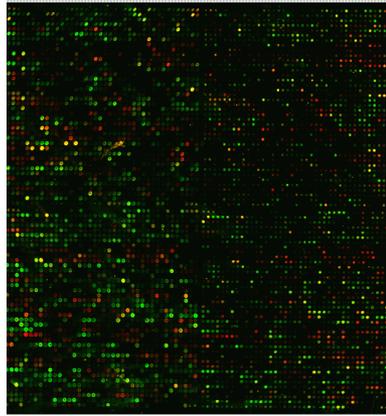


	Normale	Tumore
Gene A	1	1
Gene B	1	2
Gene C	2	1
Gene D	0	0

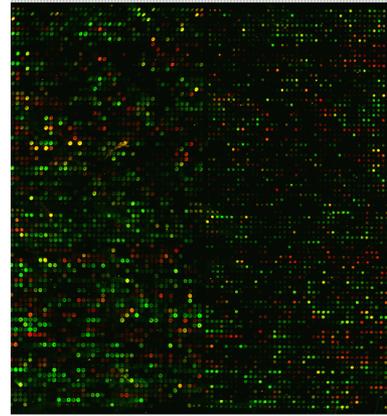
Scannerizzazione img.



Array1

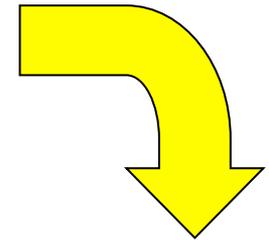


Array2



Array3

...



Arrays

	Array1	Array2	Array3	Array4	Array5	...
Gene1	0.46	0.30	0.80	1.51	0.90	...
Gene2	0.10	0.49	0.24	0.06	0.46	...
Gene3	0.15	0.74	0.04	0.10	0.20	...
Gene4	0.45	1.03	0.79	0.56	0.32	...
Gene5	0.06	1.06	1.35	1.09	1.09	...
...

rapporto di espressione del
gene 5 nell'array 4

Le “dimensioni” dell’informazione:

- Esperimento tipico: una tabella di numeri con più di 22000 righe (geni) e sessanta array (campioni).
- Se stampassimo l’intera tabella con un carattere di stampa di dimensioni editoriali (12pt) ci vorrebbero quasi 3700 fogli formato A4...
- ...pari ad una superficie totale di circa 220 metri quadri!

Analisi quantitativa

- Finalità: trovare gruppi di geni “rilevanti” nei meccanismi di insorgenza di una malattia
- Problema: si può definire il significato di “rilevante” in varie maniere:
 - Def. 1:

Geni che mostrano una differenza di espressione significativa nelle cellule tumorali rispetto a quelle normali (espressione media, distribuzione di espressione diversa, ecc...)

→ problema del confronto tipicamente statistico

Analisi quantitativa

– Def. 2:

Geni che hanno un alto potere diagnostico o prognostico,

ovvero gruppi di geni che possono essere utilizzati per determinare regole di classificazione di nuovi campioni sulla base di osservazioni pregresse

→ classico problema di *apprendimento automatico*

– Def. 3:

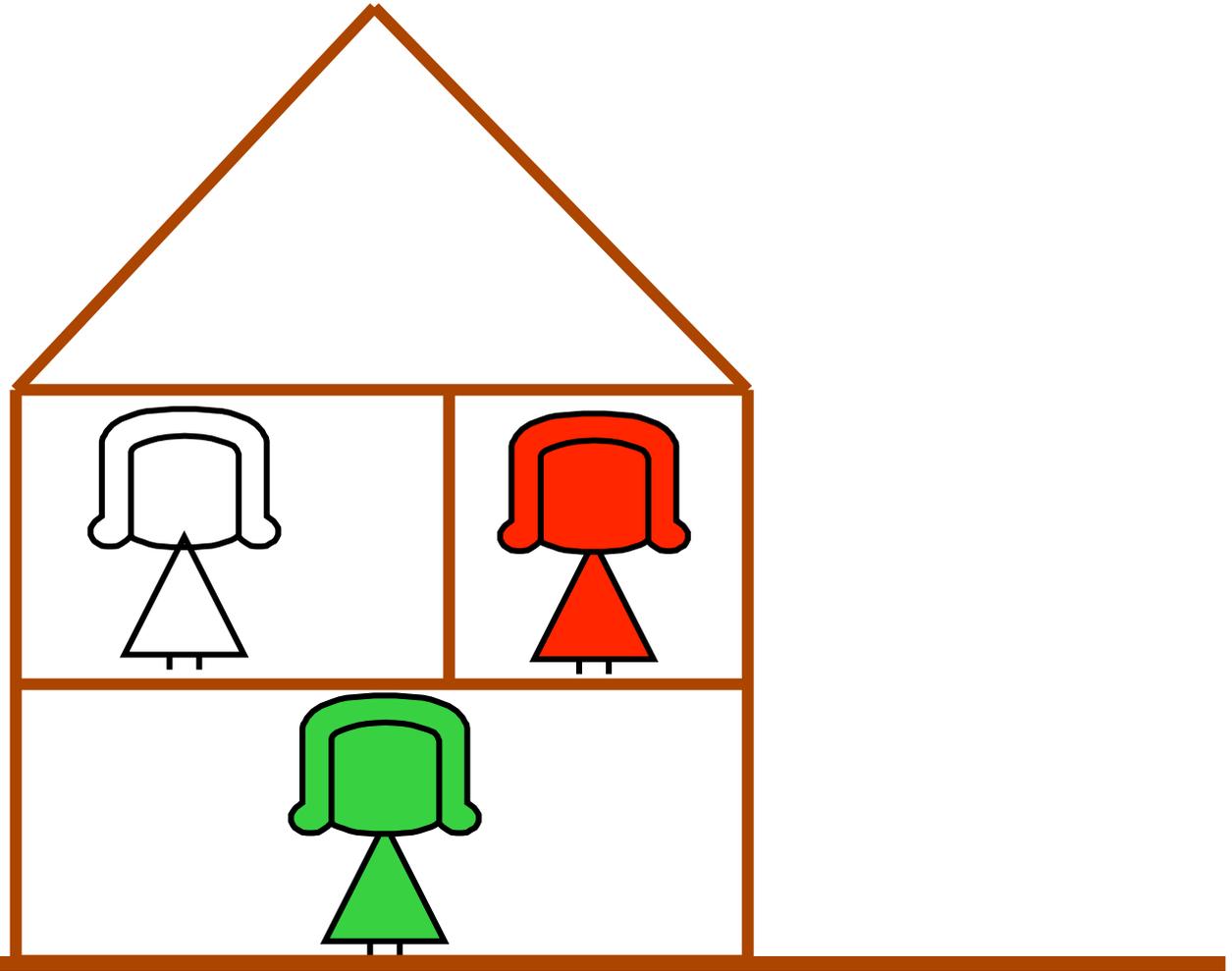
Gruppi di geni che attraverso la loro interazione determinano elevati livelli di associazione con la malattia

→ approccio con la *Teoria dei Giochi*

Analisi quantitativa

- Finalità: trovare gruppi di geni “rilevanti” nei meccanismi di insorgenza di una malattia:
- Quali sono le caratteristiche dei geni che intendo mettere in luce?
- Come valuto la rilevanza dei comportamenti che osservo in relazione a tali caratteristiche?
- L'approccio assiomatico in Teoria dei Giochi e gli indici di potere forniscono delle risposte.

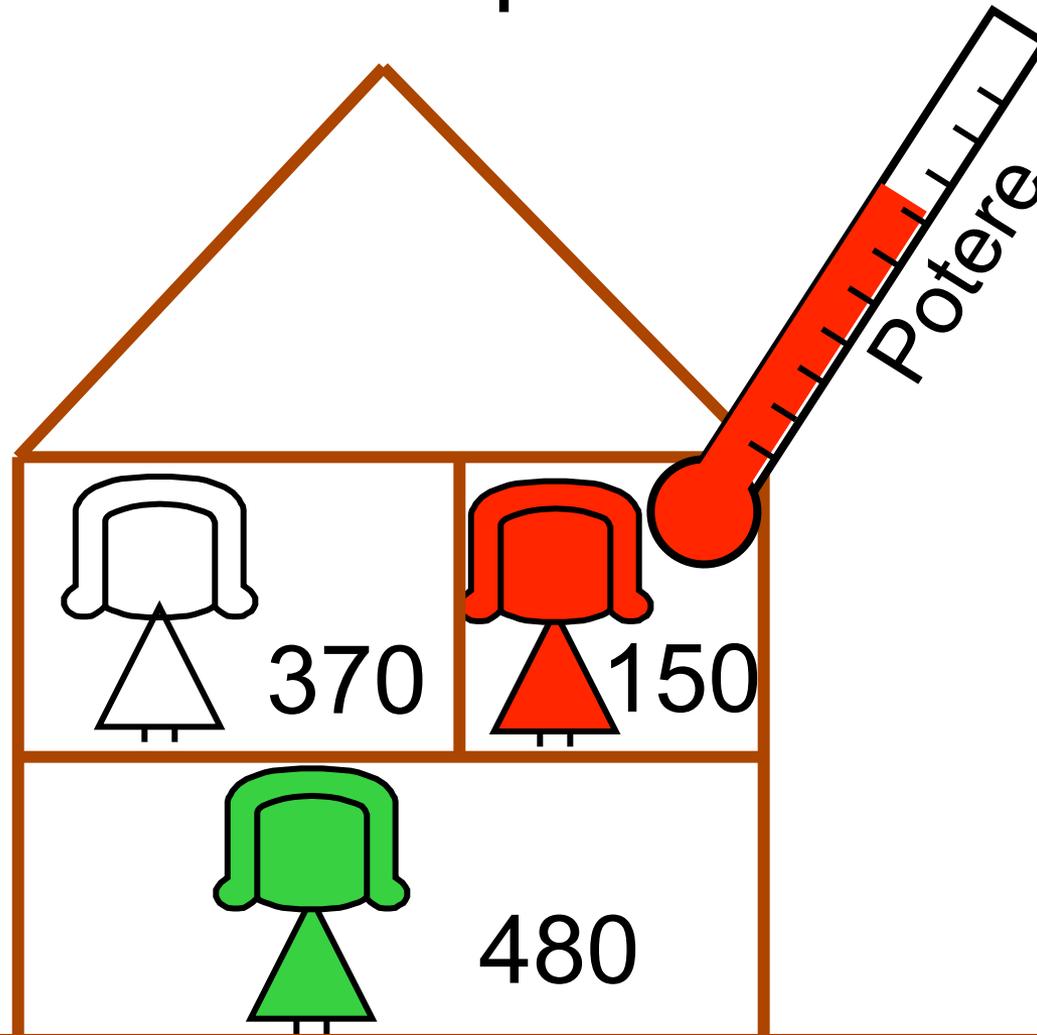
Condominio



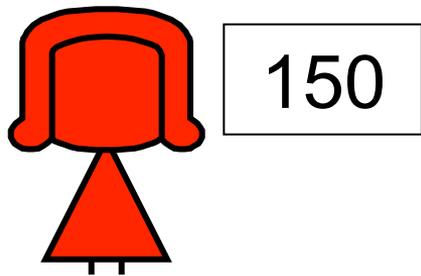
Assemblea condominiale

- **Regola decisionale:** vince il gruppo di condomini che ha più di due terzi dei millesimi.
- Come si può misurare il potere di ciascun condomino?

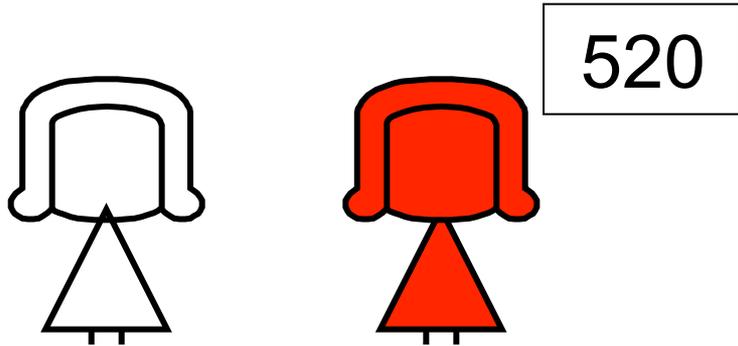
Indice di potere



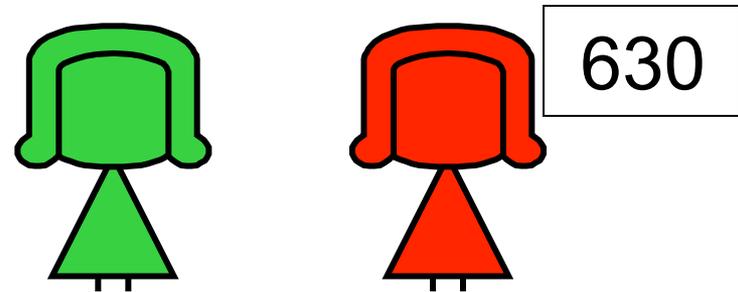
Che proprietà voglio che soddisfi un indice di potere?



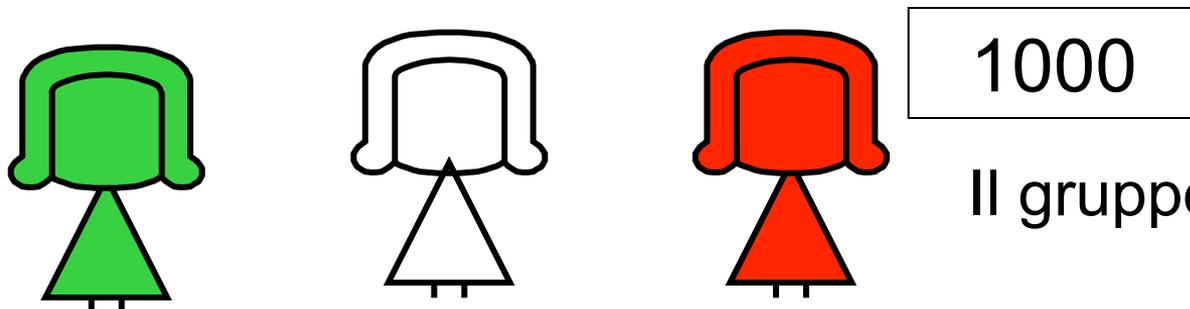
Il gruppo ha meno dei due terzi dei
millesimi



Il gruppo ha meno dei due terzi dei
millesimi



Il gruppo ha meno dei due terzi dei
millesimi

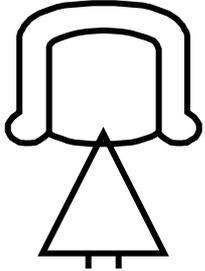


Il gruppo ha più dei due terzi dei
millesimi

0

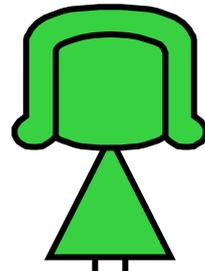
Il gruppo ha meno dei due terzi dei
millesimi

370



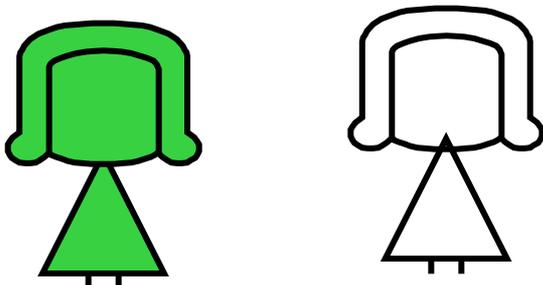
Il gruppo ha meno dei due terzi dei
millesimi

480

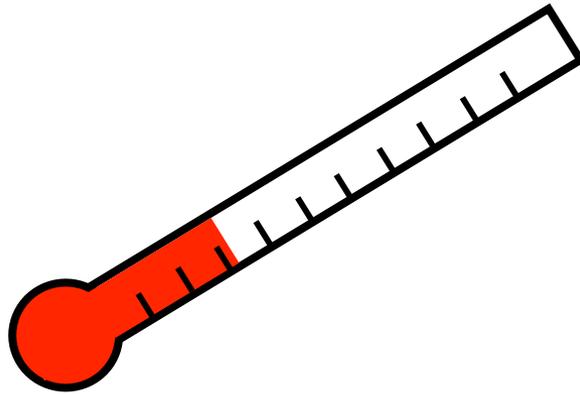
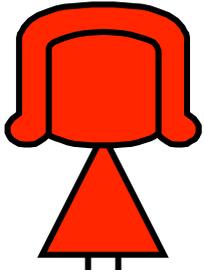


Il gruppo ha meno dei due terzi dei
millesimi

850



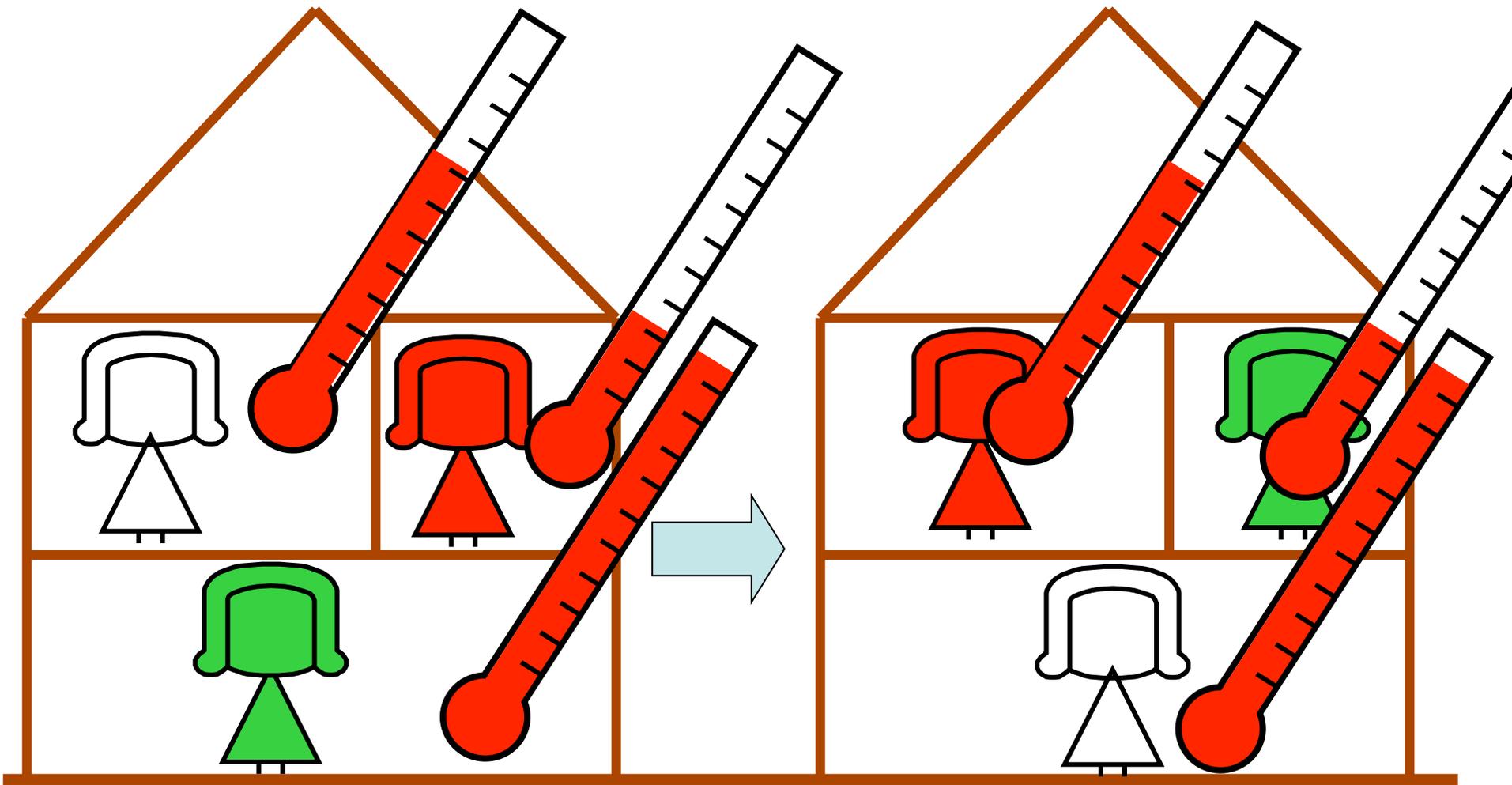
Il gruppo ha più dei due terzi dei
millesimi



= 0

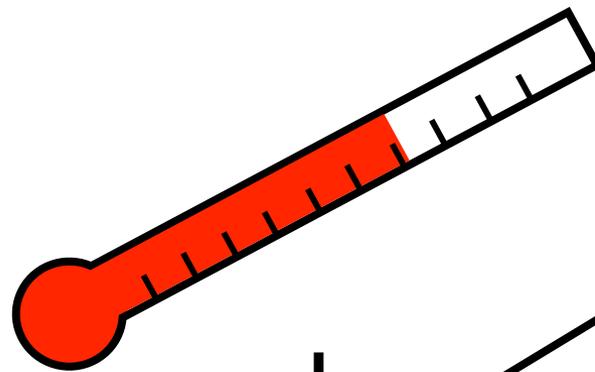
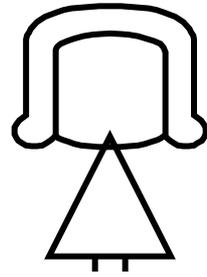
Giocatore nullo:

l'indice di potere di chi non contribuisce mai a far vincere un gruppo deve essere nullo

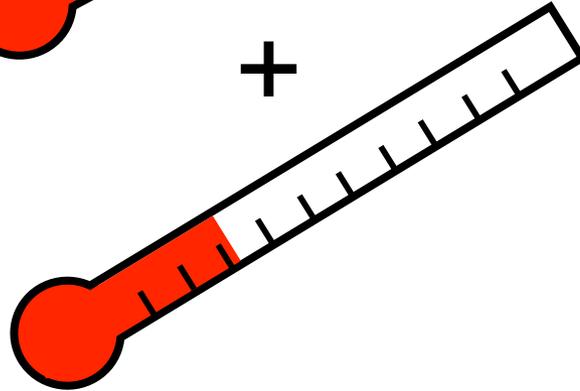


Anonimità:

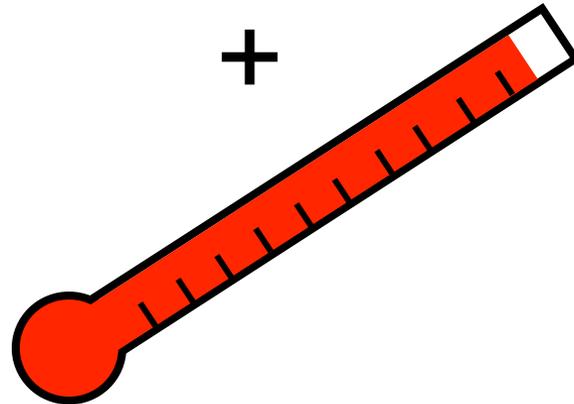
L'indice di potere non dipende dal nome dei condomini



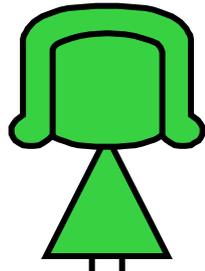
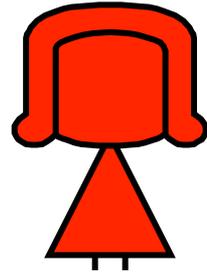
+



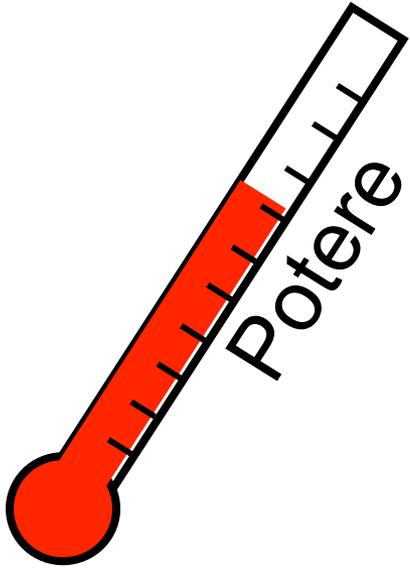
+



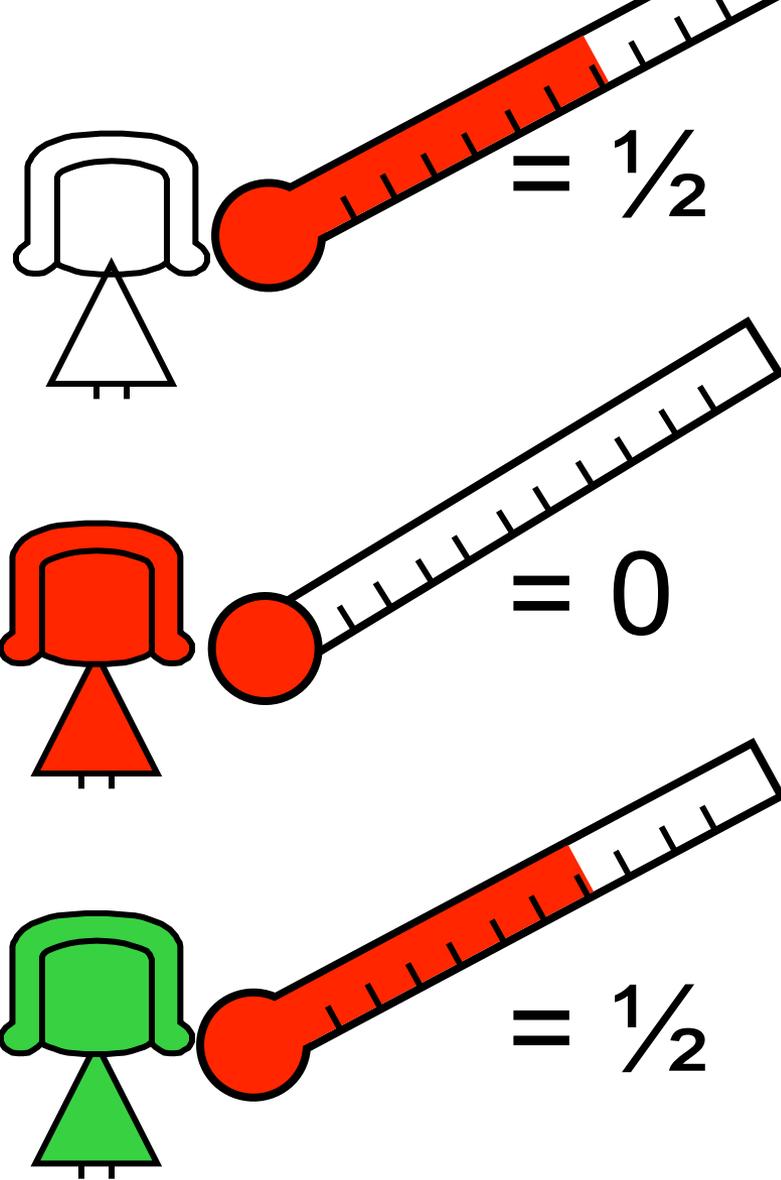
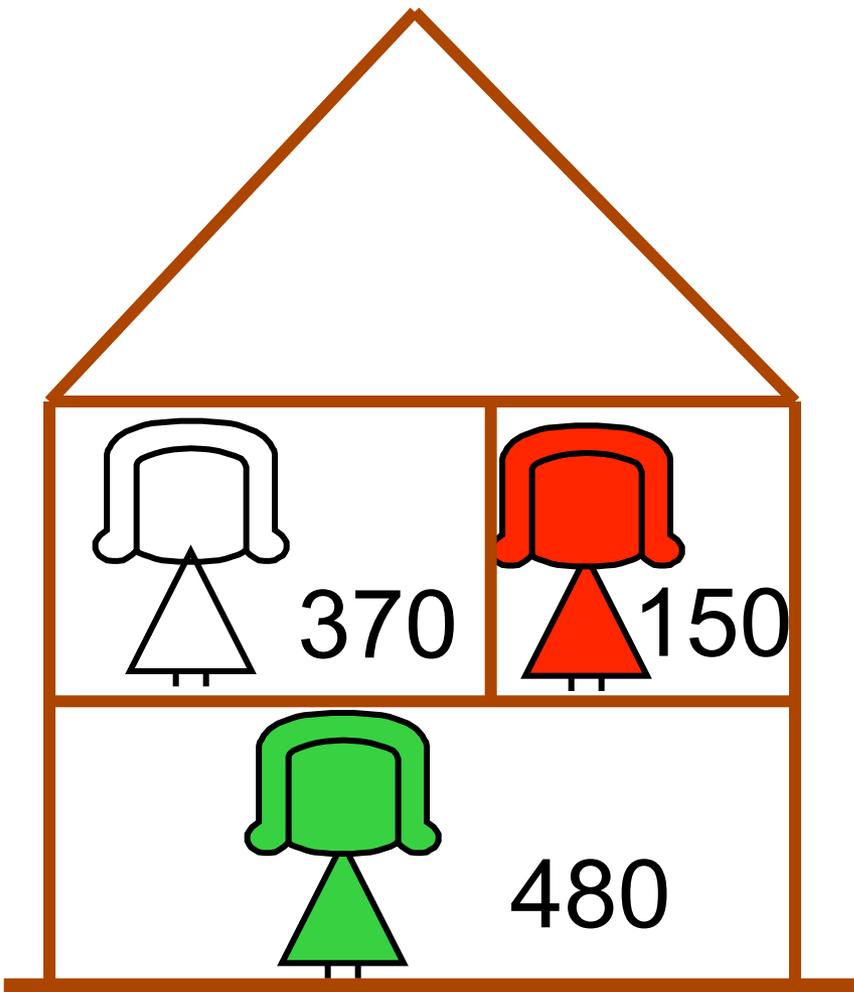
= 1



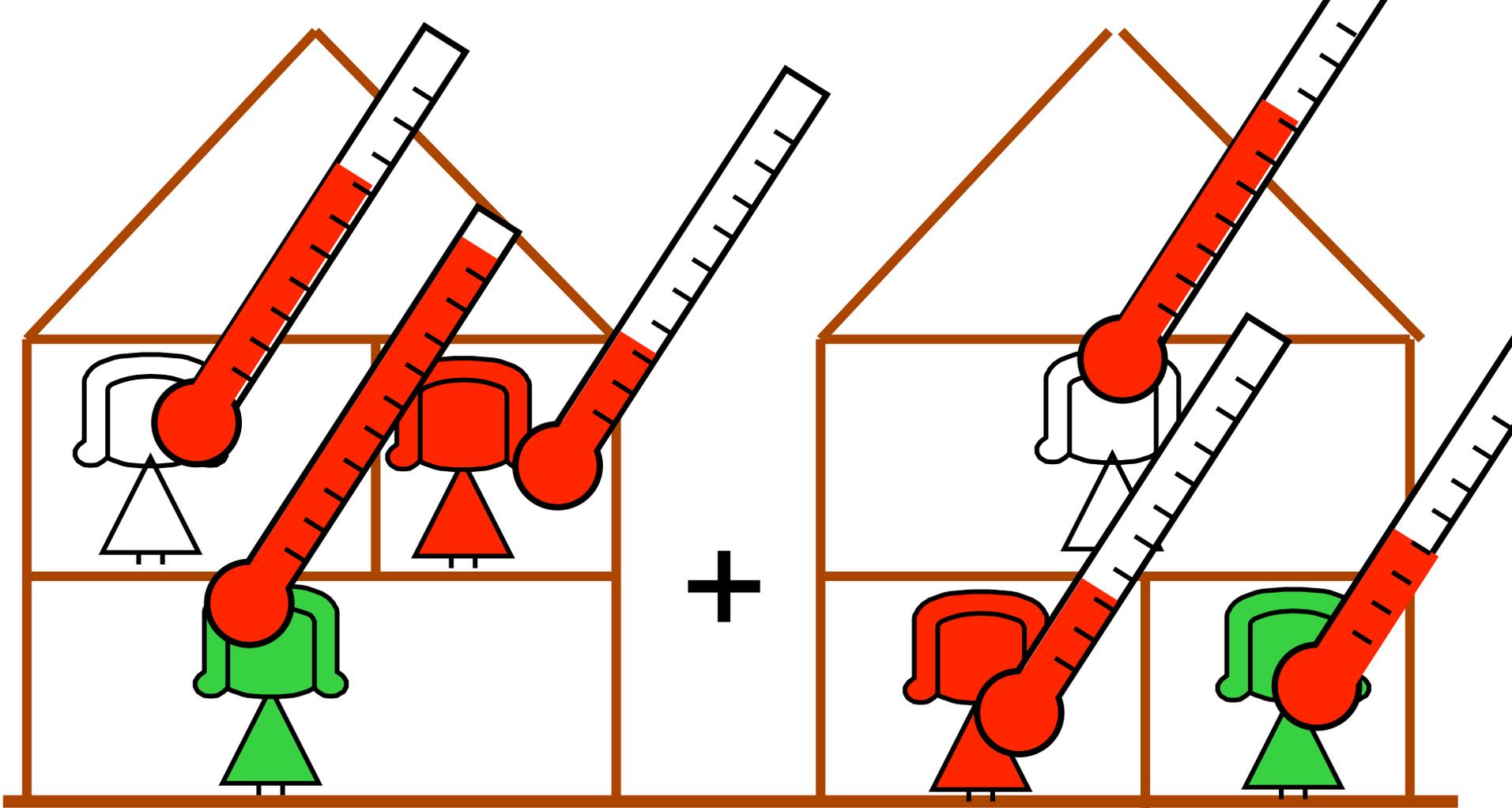
Efficienza: la somma degli indici di potere deve essere 1



Il valore Shapley (1953) soddisfa le proprietà di Anonimità, Efficienza e del Giocatore nullo...



...è l'unico che soddisfa tali proprietà (più una quarta, detta del Trasferimento) nei giochi semplici...

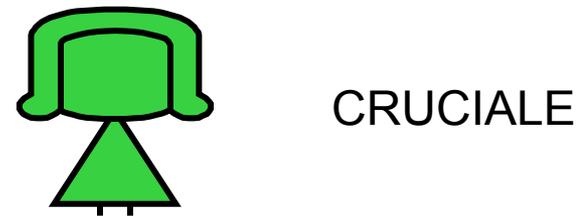
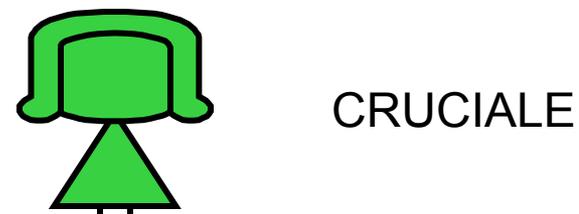
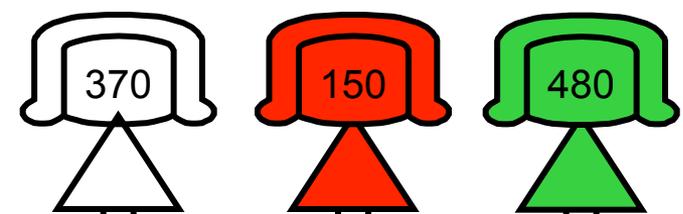
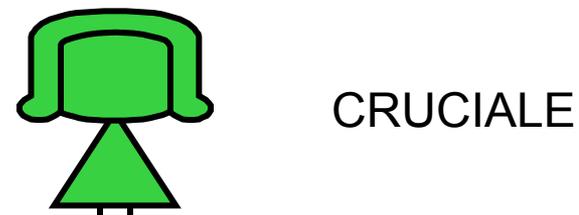
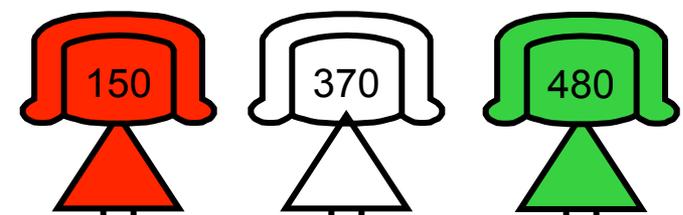
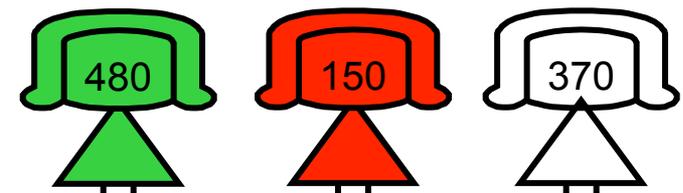
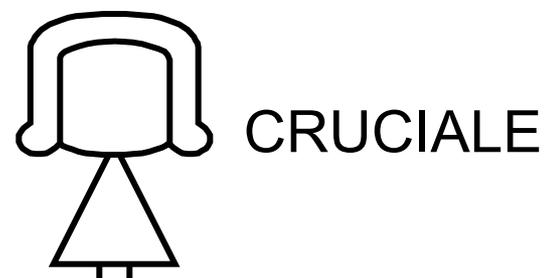
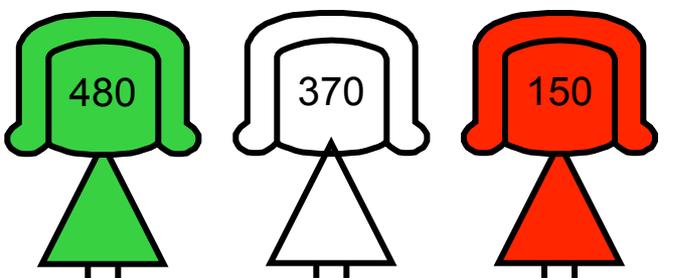


Trasferimento:

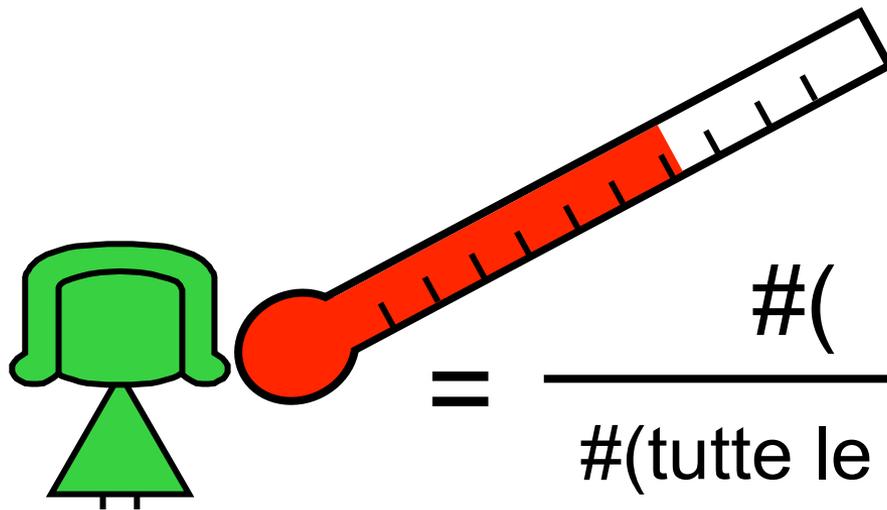
Regola il comportamento della somma di potere tra due situazioni diverse

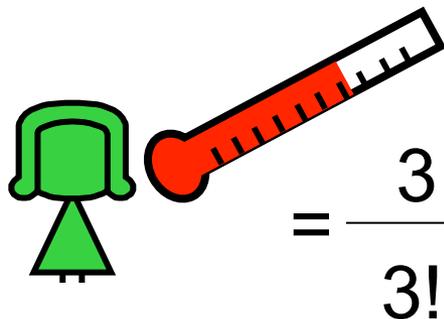
Indice di potere

- Misura della “forza” dei giocatori
- Valutazione a priori della probabilità di ciascun giocatore di giocare un “ruolo rilevante” in un processo collettivo di scelta decisionale che segua un dato insieme di regole



Shapley & Shubik (1954)


$$= \frac{\#(\text{robot cruciale})}{\#(\text{tutte le permutazioni di giocatori})}$$


$$= \frac{3}{3!} = \frac{3}{6} = \frac{1}{2}$$

Consiglio di sicurezza dell'ONU

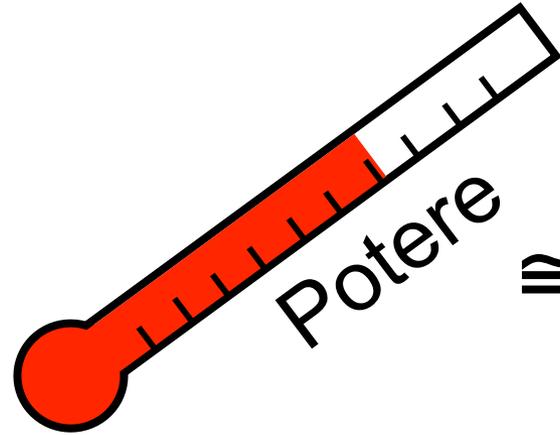
15 Nazioni membri:

- 5 **Membri permanenti**: Cina, Francia, Federazione Russa, Regno Unito di Gran Bretagna e Irlanda del Nord, Stati Uniti d'America.
- 10 **Membri non permanenti** (con mandato di due anni). Nel 2017: Bolivia, Egitto, Etiopia, Italia, Giappone, Kazakistan, Senegal, Svezia, Ucraina, Uruguay.

(<http://www.un.org/>)

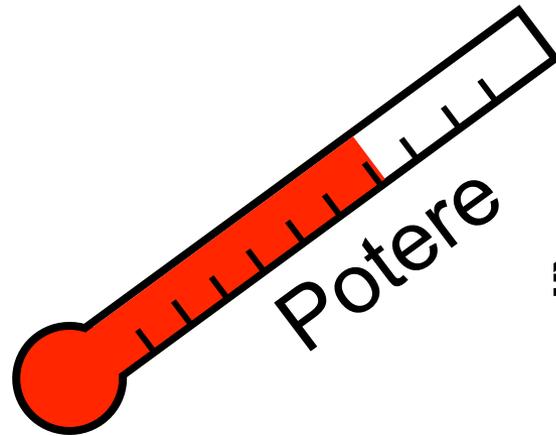
Consiglio di Sicurezza dell'ONU

- **Regola decisionale:** le **decisioni sostanziali** - come per esempio l'utilizzo di misure dirette per la risoluzione di conflitti - richiedono il voto positivo di nove membri.
 - È sufficiente il voto negativo di uno dei membri permanenti - il **veto** - per annullare la decisione.
- Quanto vale il potere di un membro non permanente all'interno del Consiglio di Sicurezza in materia di decisioni sostanziali?



Potere

≈ 19.6%



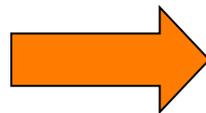
Potere

≈ 0.2%

Scendono in campo i geni ...

- I giocatori sono proprio i **geni**
- Ma chi fornisce la **regola decisionale** nel contesto dei geni?
- Possibile risposta: costruiamo una regola decisionale sulla base dei dati di espressione genica
- Definiamo un criterio per stabilire quali geni si comportano in maniera “anormale” su ciascun array.

	array1
gene1	1.121
gene2	2.453
gene3	3.586



	array1
gene1	0
gene2	1
gene3	1

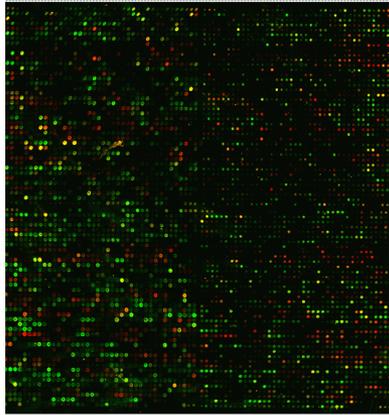
Regola decisionale

Un gruppo di geni è “*vincente*” in un singolo esperimento se **tutti i geni** che si comportano in maniera anormale nell’ esperimento sono **contenuti** in quel gruppo.

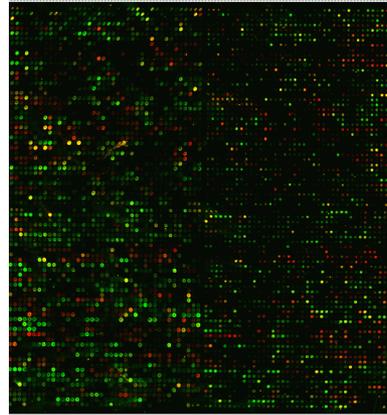
	array1
gene1	0
gene2	1
gene3	1

Esempio:

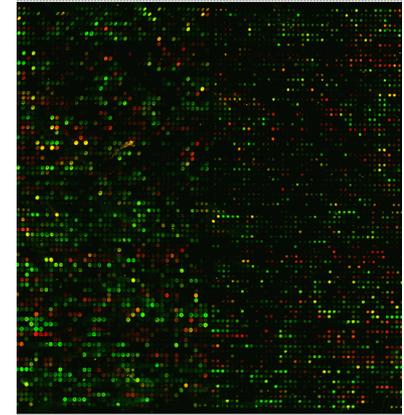
Il gruppo {gene2, gene3} e il gruppo {gene1, gene2, gene3} sono entrambi vincenti.



Array1



Array2



Array3

...

	array1	array2	array3
gene1	0	1	0
gene2	1	1	0
gene3	1	0	1

- Il gruppo {gene2, gene3} è vincente due volte su tre;
- Il gruppo {gene1, gene2} è vincente una volta su tre
- Così via per tutti i possibili gruppi...

Esempio:

	Array1	Array2	Array3
g_1	0	1	0
g_2	1	1	0
g_3	1	0	1

A questa matrice corrisponde il *microarray game* $\langle \{g_1, g_2, g_3\}, v \rangle$ tale

che

$$v(\emptyset) = v(\{g_1\}) = v(\{g_2\}) = 0$$

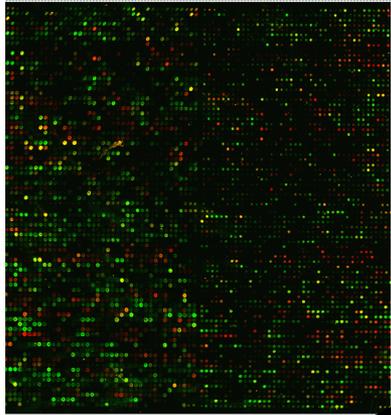
$$v(\{g_1, g_3\}) = v(\{g_1, g_2\}) = v(\{g_3\}) = 1/3$$

$$v(\{g_2, g_3\}) = 2/3$$

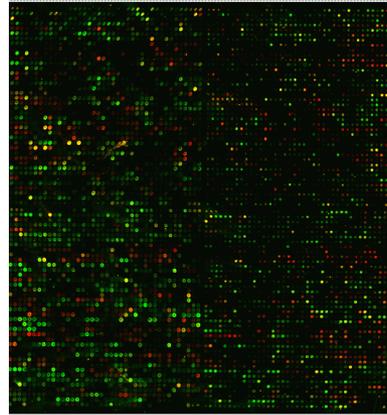
$$v(\{g_1, g_2, g_3\}) = 1.$$

Il valore Shapley è:

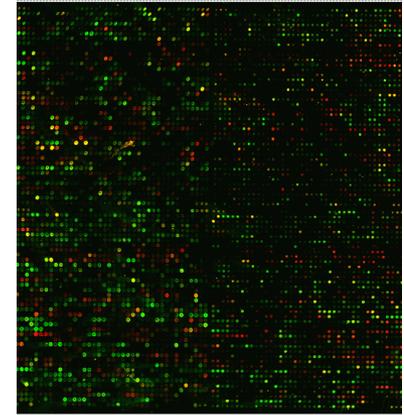
$$Sh_{g_1} = 1/6 \quad Sh_{g_2} = 1/3 \quad Sh_{g_3} = 1/2$$



Array1

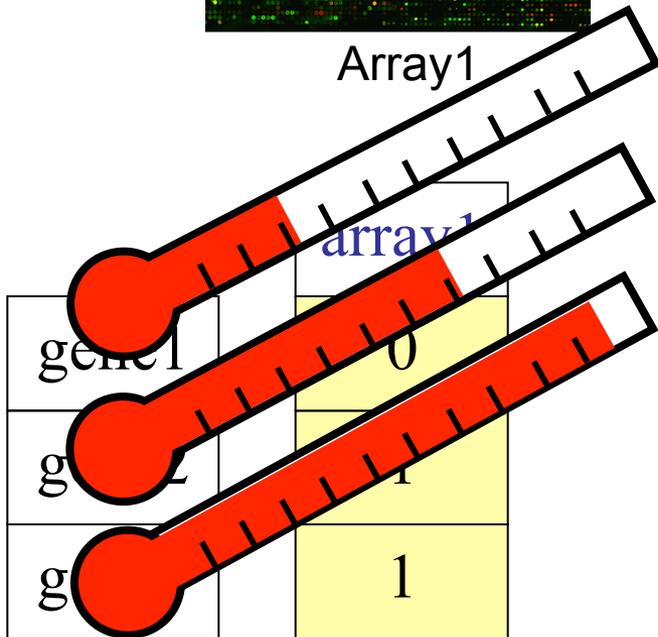


Array2



Array3

...



array2
1
1
0

array3
0
0
1

Possiamo utilizzare il valore Shapley per ordinare i geni in base alla loro rilevanza per l'insorgere di una malattia genetica.

Il valore Shapley come indice di rilevanza di geni

- Perché possiamo usare il valore Shapley In questo contesto?
- Approccio assiomatico: giustifichiamo l'uso del valore Shapley attraverso alcune **proprietà** che esso soddisfa
- Proprietà **con interpretazione biologica**

Gene regulatory pathway (GRP)

Un insieme di geni in una cellula che interagiscono tra loro, regolando il livello di espressione dei geni all'interno dell'insieme stesso.

Il meccanismo di regolazione dell'espressione genica in molte pathway è tutt'ora lungi dall'essere compreso a pieno.

Una delle maggiori difficoltà nella comprensione dei meccanismi che governano tali pathway è l'alto numero di geni coinvolti in un'analisi con i microarray.

Ricostruire pathway da dati di espressione genica è un passo cruciale nella comprensione della funzione di geni nell'insorgere di una malattia genetica.

Partnership di geni

Un gruppo di geni S tale che non esiste un sottoinsieme proprio (\subset) di S in grado di contribuire al cambiamento del valore del gruppo esterno ad S .

Esempio

Questi due
insieme sono
partnership di geni
nel gioco di
microarray
corrispondente

	a1	a2	a3
g1	0	1	1
g2	0	1	1
g3	1	0	1

Partnership di geni

Il concetto di partnership in TU-games è stato introdotto in Kalai e Samet (1988) in un contesto che non coinvolge geni per caratterizzare quelle coalizioni che si comportano come un individuo (tutte le sottocoalizioni non hanno alcun potere).

Motivazioni nell'adottare la definizione di partnership di geni come una buona rappresentazione di un GRP:

- Non richiede informazioni a priori sui meccanismi di regolazione tra geni
- Richiede che non esista un sottogruppo proprio di geni che interagisca con un gene o con un gruppo di geni esterni per provocare l'insorgenza della malattia.

Assiomi per il valore Shapley sui *microarray games*

Proprietà 1: Gene Nullo (NG)

un indice di rilevanza deve attribuire rilevanza nulla ai geni che non sono mai anormalmente espressi nelle cellule malate.

	a1	a1	a1
g_1	0	1	0
g_2	0	0	0
g_3	1	0	1

$$\psi_2 = 0$$

Assiomi per il valore Shapley sui *microarray games*

Proprietà 2: **Equal Splitting (ES)**

tutti gli esperimenti devono essere considerati ugualmente affidabili e quindi avere lo stesso peso nel calcolo del potere dei geni.

	a1		a2			a1	a2			
g1	0	ψ_1	+	1	ψ'_1	=	g1	0	1	$(\psi_1 + \psi'_1)/2$
g2	0	ψ_2		1	ψ'_2		g2	0	1	$(\psi_2 + \psi'_2)/2$
g3	1	ψ_3		0	ψ'_3		g3	1	0	$(\psi_3 + \psi'_3)/2$

Proprietà 3: **Monotonia delle Partnership (PM)**

Se si hanno due partnerships di geni S e T, con $|T| \geq |S|$ e che siano

- disgiunte: $S \cap T = \emptyset$,
- equivalenti: $v(S) = v(T)$ ed
- Esaustive: $v(S \cup T) = v(N)$

i geni nella *partnership* meno numerosa S devono ricevere più rilevanza di quelli in T.

Esempio

		S ₁	S ₂
ψ_1	g1	0	1
ψ_2	g2	0	1
ψ_3	g3	1	0
ψ_4	g4	1	0
ψ_5	g5	1	0

$$\psi_i \geq \psi_k$$

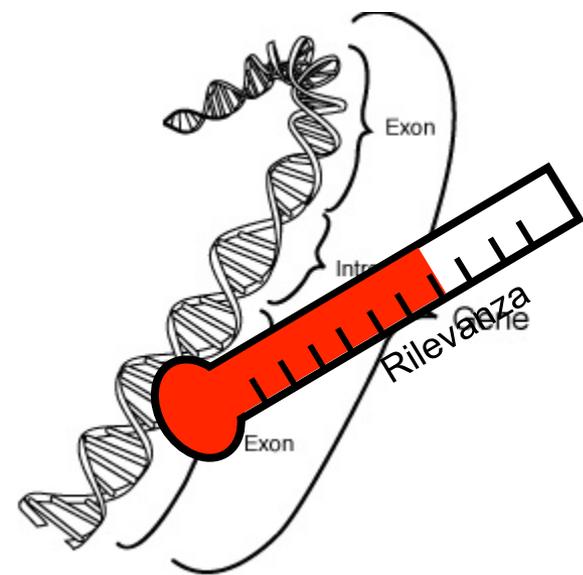
per ogni
 $i \in \{1, 2\}$
 $k \in \{3, 4, 5\}$

Proprietà 4: **Razionalità di *partnership* (PR)**

Il valore totale di rilevanza ricevuta da una *partnership* S dovrebbe essere non inferiore a $v(S)$

Proprietà 5: **Fattibilità di *partnership* (PF)**

Il valore totale di rilevanza ricevuta da una *partnership* S dovrebbe essere non superiore a $v(N)$



Teorema (Moretti, Patrone, Bonassi (2007)):
Il valore Shapley è l'unico indice che soddisfa le proprietà NP, ES, PM, PR, PF sulla classe dei giochi di *microarray*.

Methodology article

Open Access

Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution

Stefano Moretti^{*1}, Danitsja van Leeuwen², Hans Gmuender³, Stefano Bonassi⁴, Joost van Delft², Jos Kleinjans², Fioravante Patrone⁵ and Domenico Franco Merlo¹

Address: ¹Epidemiology and Biostatistics, National Cancer Research Institute, Genova, Italy, ²Department of Health Risk and Toxicology, Maastricht University, Maastricht, the Netherlands, ³Canoplate AG, Basel, Switzerland, ⁴Unit of Molecular Epidemiology, National Cancer Research Institute, Genova, Italy and ⁵DIPTM, University of Turin, Italy

Email: Stefano Moretti^{*} - stefano.moretti@isg.unige.it
Hans Gmuender - hans.gmuender@genedata.com
Jos Kleinjans - J.Kleinjans@GRAT.unimaas.nl

^{*} Corresponding author

Published: 2 September 2008

BMC Bioinformatics 2008, 9:361 doi:10.1186/1471-2108-9-361

This article is available from: <http://www.biomedcentral.com/10.1186/1471-2108-9-361>

© 2008 Moretti et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Cancer

Original Article

Identification of low intratumoral gene expression heterogeneity in neuroblastic tumors by genome-wide expression analysis and game theory[†]

Domenico Albino MS^{1,†}, Paola Scaruffi PhD^{2,†}, Stefano Moretti PhD³, Simona Coco PhD², Mauro Truini MD⁴, Claudio Di Cristofano MD⁵, Andrea Cavazzana MD⁶, Sara Stigliani PhD², Stefano Bonassi PhD⁷, Gian Paolo Tonini PhD^{2,§,*}

Issue



Cancer

Volume 113, Issue 6, pages
1412–1422, 15 September
2008

Article first published online: 31 JUL 2008

DOI: 10.1002/cncr.23720

Copyright © 2008 American Cancer Society

Esercizio:

	sample1	sample2	sample3	sample4
gene1	0	0	1	0
gene2	1	0	0	1
gene3	0	1	0	0
gene4	1	0	0	0
gene5	1	0	0	0
gene6	0	1	1	0
gene7	0	0	1	0
gene8	1	0	0	1
gene9	0	0	0	0
gene10	1	0	1	1

•Calcolare il valore Shapley del gioco di microarray associato a questa tabella.

•La coalizione {gene2, gene3, gene 4} è una partnership?

•La coalizione {gene2, gene 8} è una partnership?

N.B. E' semplice calcolare il valore Shapley poiché un microarray game può essere scritto come combinazione lineare di giochi di unanimità, uno per ogni array e con un coefficiente pari al numero di campioni.

Bibliografia

-  D.Albino et al. "Identification of low intratumoral gene expression heterogeneity in neuroblastic tumors by genome-wide expression analysis and game theory." *Cancer*, 113.6: 1412-1422 (2008).
-  S. Moretti, F. Patrone: "Transversality of the Shapley value." *Top*, 16:1-41 (2008).
-  S. Moretti, F. Patrone, S. Bonassi: "The class of microarray games and the relevance index for genes." *Top*, 15:256-280 (2007).

Bibliografia

-  S. Moretti et al. "Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution." BMC bioinformatics, 9.1: 361 (2008).
-  L. S. Shapley: "A Value for n-Persons Games." Contributions to the Theory of Games. Vol. II Annals of Mathematical Studies, 28: 307-317 Princeton University Press, Princeton (1953).
-  L. S. Shapley, M. Shubik: "A Method for evaluating the distribution of power in a committee system." American Political Science Review, 48:787-792 (1954).