

lezione 13-14

martedì 24 Novembre 2009

corso di genomica

a.a. 2009/10

aula 6a ore 14.00-16.00

corso di laurea specialistica
magistrale Biotecnologia

lezione 11 Dicembre sequenziamento shot-gun metodo
pyrofosfato 454 e 480 Roche. Dr.Rodriguez

lezione 15 Dicembre Programmi informatici per confronti
genomici. Dr.P. Daddabbo

limiti e pregiuduzi con WGA

- identificate centinaia di varianti genetiche di malattie complesse
- nuove conoscenze di architetture genetiche (interazioni)
- identificazioni di varianti con piccoli incrementi di rischio
- spiegano solo piccola parte della familiarità
- restano molte domande sulla spiegazione delle rimanenti componenti genetiche non trovate o mancanti
- ci sono altre strategie oltre il WGA per spiegare la genetica delle malattie complesse

familiarità nelle malattie comuni

studi precedenti prima dei WGA studies :

familiarità

geni candidati

WGA linkage studies di malattie complesse non ha successo su per il basso potere di risoluzione per varianti con effetto modesto (confuso col background)

il razionale del WGA = malattia comune-varianti comuni attribuibili ad alleli con frequenza $> 1-5\%$

possibile per i milioni di SNPs identificati su “chips” commerciali

esempi di architettura allelica

molte strutture alleliche hanno funzione ancora ignota

degenerazione della macula (parte della retina) con l'età
dovuta a:

- varianti a largo effetto (raddoppio del rischio)
- varianti comuni singole o in combinazione danno aumento di rischio di 1.1 - 1.5 volte

spiegano solo piccola parte dell'ereditarietà

oltre 40 loci attribuiti all'altezza con ereditarietà oltre al 80%
spiegano solo 5% della varianza fenotipica con studi su
decine di migliaia di soggetti

superamento del bias su regioni codificanti / non codificanti
(di cui poco si sapeva e si sa)

inefficienza del WGA ?

perchè la gran parte dei fattori ereditari non sono identificati?

- alta proporzione delle differenze individuali per la suscettibilità a malattie dipende da fattori genetici polimorfici

- capire questa variabilità genetica utile alla diagnosi, trattamento e prevenzione

- molte aspettative e qualche delusione

- primi studi di WGA: fenotipizzazioni imprecise, controlli con comparabilità dubbia hanno ridotto le stime degli effetti

- comunque hanno superato le aspettative identificando molti tratti genetici, che però spiegano solo una piccola proporzione della ereditarietà stimata della patologia

varie possibilità di inadeguatezza

- maggior numero di varianti con piccolo effetto,
 - varianti più rare non identificate per la bassa frequenza
 - varianti strutturali poco presenti dagli arrays esistenti
 - poca forza nel determinare interazione gene-gene
 - inadeguatezza nel tener conto dell'ambiente condiviso coi parenti
-
- esistenza di materiale genetico inidentificato (materia scura) invisibile di cui si conosce l'effetto. Cioè la somma degli elementi scoperti copre solo in parte la componente genetica di malattie complesse.

allelic architecture

architettura allelica : numero, tipo, effetto di grandezza, frequenza

l'effetto genetico mancante potrebbe avere forma diversa nelle varie patologie al momento attuale non identificabile

nei caratteri complessi appunto può cambiare l'architettura genetica degli alleli che si combinano per tutte le caratteristiche elencate sopra, in cui ognuna di esse può avere un peso diverso

da qui l'effetto può diventare quantitativo (vedi lez. 13-14)

esempio calzante di malattia comune

degenerazione della macula per anzianità:

- ereditarietà spiegata da piccolo numero di varianti comuni con forte effetto, le stesse in altre patologie come Crohn disease hanno un peso di ereditarietà minore benchè si sia trovato un numero maggiore di varianti.

difficile determinare le differenze dei tratti di architettura genetica come predetto dalla clinica ed epidemiologia.

alcune differenze apparenti possono dipendere semplicemente dalle conoscenze sui caratteri stessi

studi in diverse condizioni dimostrano che il n. di varianti aumenta con l'ampiezza del campione (esono gli alleli rari)

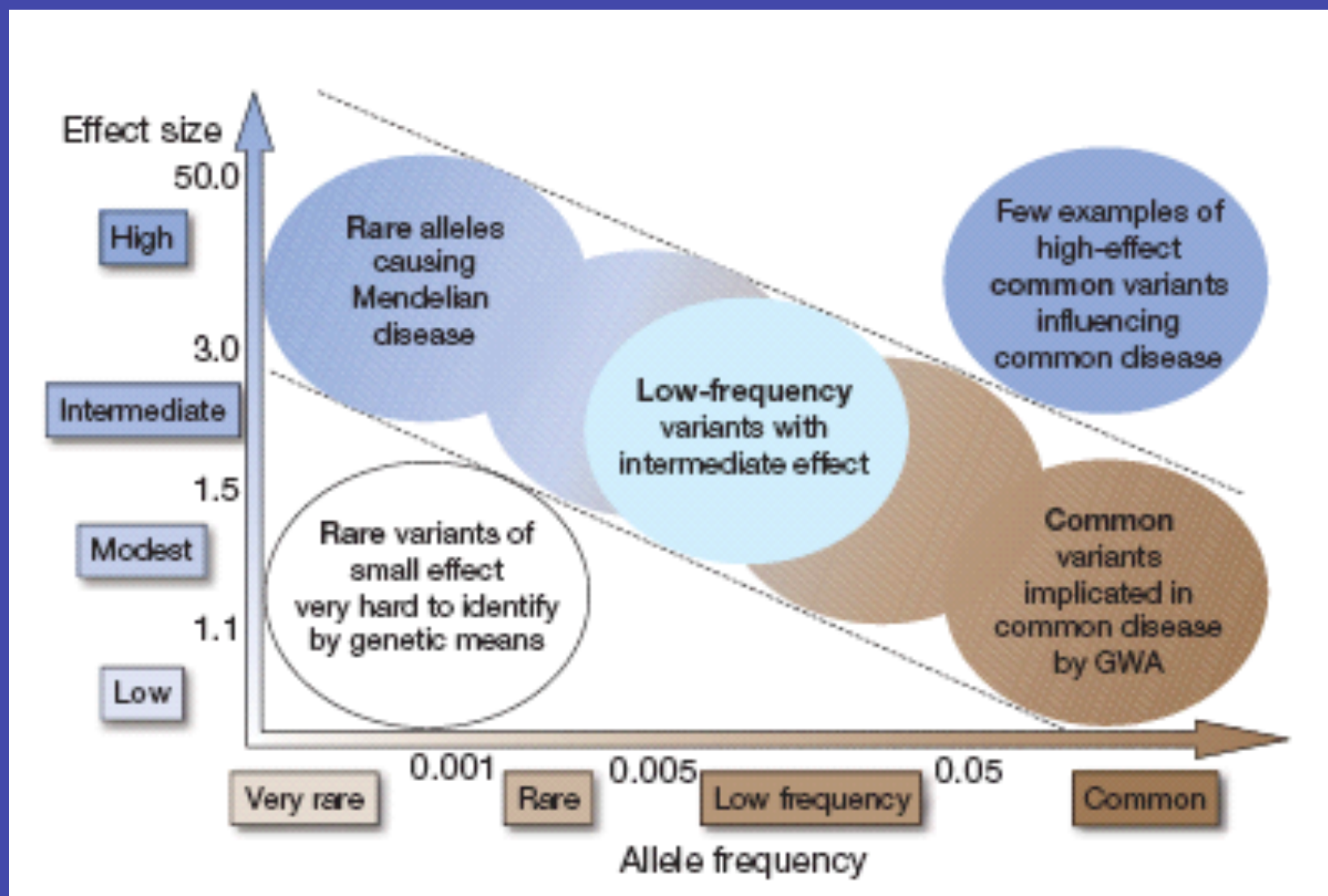
stima dell'ereditarietà e n di loci di tratti complessi

malattia	n.loci	proporzione*	metodo
Age-related macular degeneration72	5	50%	sibling rec.risk
Crohn's disease21	32	20%	genetic risk**
Systemic lupus erythematosus	6	15%	sibling rec.risk
Type 2 diabetes	18	6%	sibling rec.risk
HDL cholesterol	7	5.2%	residual phenot.var.
Height	40	5%	phenot.variance
Early onset myocardial infarction	9	2.8%	phenot.variance
Fasting glucose	4	1.5%	phenot.variance

*proporzione di ereditabilità spiegata dallo studio

** liability predisposizione

allelic freq vs. effect size



legend to figure

Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

Probabilità di identificare varianti genetiche con la frequenza di rischio allelico e forza dell'effetto genetico (odds ratio). L'interesse e l'enfasi sta nell'identificare associazioni tra le caratteristiche mostrate nella diagonale a puntini.

T.A.Manolio et al. Nature Reviews vol.461 October 2009, 747-753. Finding the missing heritability of complex diseases.

scarsa di varianti e pop. genet.

- numero basso di varianti genetiche con predisposizione di larghe proporzioni collegato alla bassa fitness
 - la fitness riduce le varianti ad alto rischio di malattia
- tipico delle malattie neuropsichiatriche (autismo ecc.) bassa riproduttività
- la riproduttività al presente non rispecchia e non correla con le forze selettive presenti durante l'evoluzione umana
- riduttivo considerare gli effetti su una singola malattia ignora gli effetti pleiotropici di delle varianti che condizionano simultaneamente altre funzioni coinvolte in altre funzioni

fattori selettivi

stabilizzazione a lungo termine elimina individui con tratti estremi

- riduzione di effetti di alleli già presenti o eliminando

mutazioni de-novo

- se un fattore partecipa ad altre malattie con selezione forte

- se fattori ambientali possono influenzare l'intensità della malattia

- malattie immunomediate (Crohn's) con effetto pleiotropico con la risposta immunitaria riducono l'effetto selettivo sulle varianti all.

- forti cambiamenti ambientali moderni possono mostrare il rischio di queste varianti associato alla malattia

- la selezione può spiegare le frequenze alleliche delle malattie e gli effetti allelici sono bassi, ma pur bassi non dovrebbero essere assenti !

statistica applicata al WGA

Odds ratio (OR): In genetic studies looking for associations between SNPs and diseases, including GWA studies, the odds ratio (OR) is the statistic conventionally used to compare the frequency of SNP alleles between the Case and Control groups in order to determine whether a given SNP is “associated” with the disease, or not. The odds ratio is a way of comparing whether the probability of a certain event is the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. When an odds ratio is greater than one when comparing cases to controls, it implies that, for a given SNP, one of the alleles is more likely to be found in the case group than in the controls. In many studies reported in the scientific literature, ORs are calculated on either the difference in allele frequencies for a given SNP between the cases and controls (allele-based single-locus test), or by observing the number of homozygotes (for each allele) and heterozygotes in the cases and controls (genotype-based single-locus test). For example, using an allele-based single-locus test and a SNP that detects alleles M and m, an OR of 1.3 would mean that allele M is 1.3 times more likely to be carried by an affected individual than allele m (assuming multiplicative disease risks and Hardy-Weinberg equilibrium at the SNP in cases and controls). For studies which report an OR for genotype MM relative to mm, an OR of 1.3 would mean that an affected individual is 1.3 times more likely to have marker genotype MM than mm. In GWA studies, a more sophisticated statistical model is often used (e.g., logistic regression) which can consider multiple SNPs, environmental factors and their interactions simultaneously. Most SNPs associated with diseases and conditions have relatively modest effects ($ORs < 2$) so an $OR > 1$ should be viewed more as an indicator of a high likelihood of a real association between the SNP and disease or condition, rather than a strict estimate of the relative risk of the disease. The odds ratio is often used as an approximation of relative risk although, strictly speaking, they are not identical.

penetranza penetrance

Penetrance: How genotype dictates disease status: the percentage of individuals with a specific genotype that possess an associated phenotype. For example, some alleles are highly penetrant, meaning that nearly all individuals with such an allele will exhibit the trait of interest. Such alleles are more easily detected in genetic studies compared to alleles of reduced or low penetrance since in these cases it is difficult to distinguish genetic from environmental factors (in other words, many people will have such an allele but will not exhibit symptoms of the disease). Penetrance of a disease might also be age-related.

fenotipo multifattoriale, poligenico

Polygenic disease: Such diseases are characterized by a clear-cut hereditary component (i.e. the disease aggregates in families but without the usual Mendelian inheritance ratios for monogenic disease). The complexity of these diseases arises from the fact that a given disease may be influenced by a relatively small number or several hundred genes. Each gene might contribute an equal but small effect, or alternatively a disproportionate effect, to the susceptibility and pathology of the disease. One of the major aims of genome wide association studies is to elucidate the identity and involvement of all the genes in the molecular pathology of a polygenic disease.

o.r. e rischio

The results are typically reported in odds-ratios, that is the ratio between the fraction (probability) with the risk variant (carriers) versus the non-risk variant (non-carriers) in the groups of affected versus the controls, i.e. expressed in terms of probabilities conditional on the affection status:

$$OR = (\Pr(c|A)/\Pr(nc|A)) / (\Pr(c|C)/\Pr(nc|C))$$

pr = probability

c = carrier

nc = non carrier

A = affected (non affettato)

C = control

rischio e dominante/recessivo

In summary, the calculations (see appendix) show that for the dominant and the recessive models, where we have a risk variant carrier, “c”, and a non-carrier, “nc”, the odds-ratio of individuals is the same as the risk-ratio between these variants:

$$OR = \Pr(A|c)/\Pr(A|nc) = r$$

likewise for the multiplicative model, where the risk is the product of the risk associated with the two allele copies, the allelic odds-ratio equals the risk factor:

$$OR = \Pr(A|aa)/\Pr(A|ab) = \Pr(A|ab)/\Pr(A|bb) = r$$

Here “a” denotes the risk allele and “b” the non-risk allele. The factor “r” is therefore the relative risk between the allele types.

pr probability, Affected,

esempio di rischio composto

As an example, an individual who has the following genotypes at 4 markers associated with risk of type-2 diabetes along with the risk relative to the population at each marker (RR general population):

Chromo 3	PPARG	CC	RR(CC) = 1.03
Chromo 6	CDKAL1	GG	RR(GG) = 1.30
Chromo 9	CDKN2A	AG	RR(AG) = 0.88
Chromo 11	TCF7L2	TT	RR(TT) = 1.54

Combined, the overall risk relative to the population for this individual is: $1.03 \times 1.30 \times 0.88 \times 1.54 = 1.81$

varianti multiple

ulteriore complicazione: molte varianti in una stessa regione
con effetti indipendenti

le varianti sono state studiate individualmente
se si mettono insieme ?

- genera aumento del MAF (minor allele frequency), si riduce il
numero di analisi da fare, ma aumentano incognite:

- quale della moltitudine di varianti è responsabile x il fenotipo?
solo di pochi alleli si conosce anche la funzione associata.

GWA x lo + è servita per determinare un rischio associato e
meno per il spiegare la causa biologica