

lezione 17-18

venerdì 27 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

varianti multiple

ulteriore complicazione: - molte varianti alleliche in una stessa regione hanno effetti indipendenti

- le varianti sono state studiate individualmente

- se si mettono insieme ? si usano meno markers = <analisi

- 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

cosa si vede effettivamente

WGA: aumentare il numero di markers per non perdere regioni con informazioni importanti

equilibrio per evitare marcatori ridondanti

se in una regione cascano più marcatori di fenomeni diversi nel WGA se ne usa uno per tutti, però dovrà seguire un'analisi dei marcatori per vedere quale è il vero responsabile del fenotipo associato

WGA usa marcatori scelti ma non sempre sono gli indicatori del fenotipo associato, possono solo indicare la regione genetica associata (aplotipo sottostante)

stima delle differenze

tra individui circa 0.4% su 3×10^9

come studiare le differenze o uguaglianze (matching)
in un confronto il background genetico pesa ed in nuove
popolazioni africane moltissimo per i fattori ancestrali origin.

ridurre il n. di varianti in caso - controllo:

- aumentare il confronto di regione cromosomica
specifica sul genoma
- selezionare alleli correlati e regioni nella popolazione
di confronto (separare il background)

= diminuzione del numero di differenze tra casi dovute
ad alleli accidentali

associazioni con rari CNVs e CNPs comuni

Table 2 | Selected disease associations with rare CNVs and common CNPs

Disease	Locus	Type of CNV	Size (kb)	Population frequency	Case frequency	Effect size (OR)
Rare CNVs						
Autism/IMR ⁵⁹	16p11.2	De novo deletion	600	1×10^{-4}	1%	100
Autism ⁶⁹	16p11.2	Rare duplication	600	3×10^{-4}	0.50%	16
Schizophrenia ^{60,70}	1q21.1	Rare deletion	1,400	2×10^{-4}	0.30%	15
IMR ⁷⁹	1q21.1	Rare deletion	1,400	2×10^{-4}	0.47%	Not observed in 4,737 controls
Schizophrenia ^{60,70}	15q13.3	Rare deletion	1,600	2×10^{-4}	0.20%	12
Epilepsy ⁸⁰	15q13.3	Rare deletion	1,600	2×10^{-4}	1.0%	Not observed in 3,699 controls
IMR ^{79,81}	15q13.3	Rare deletion	1,600	2×10^{-4}	0.30%	Not observed in 960 controls
Schizophrenia ⁸²	22q11.2	Rare deletion	3,000	2.5×10^{-4}	1%	40
Common CNPs						
Crohn's disease ⁸³	IRGM	Deletion polymorphism	20	7%	11%	1.5
Body mass index ⁶¹	NEGR1	Deletion polymorphism	45	65%	Quantitative trait	<1 kg
Psoriasis ⁸⁴	LCE3C	Deletion polymorphism	30	55%	65%	1.3

IMR, idiopathic mental retardation.

copy number variation
copy number polymorphism

ca. 400 WGA publicati

underlying genetic architecture of complex traits and the predominance of non-coding variants that may have a role in their aetiology. Just as linkage studies demonstrated that complex diseases cannot be explained by a small number of rare variants with large effects, GWAS have shown that they cannot be explained by a limited number of common variants of moderate effect (Fig. 1)

- realistic effect sizes, for detecting associations with low frequency variants by GWAS

Low frequency variants of intermediate effect might also contribute to explaining missing heritability that should be tractable through large meta-analyses and/or imputation of genomewide association data.

Information on lower frequency alleles emerging from projects such as the 1,000 Genomes will be used to produce even more comprehensive GWA arrays, and will facilitate the investigation of the lower frequency spectrum without the need for de novo sequencing.

oltre i 400 studi di WGA

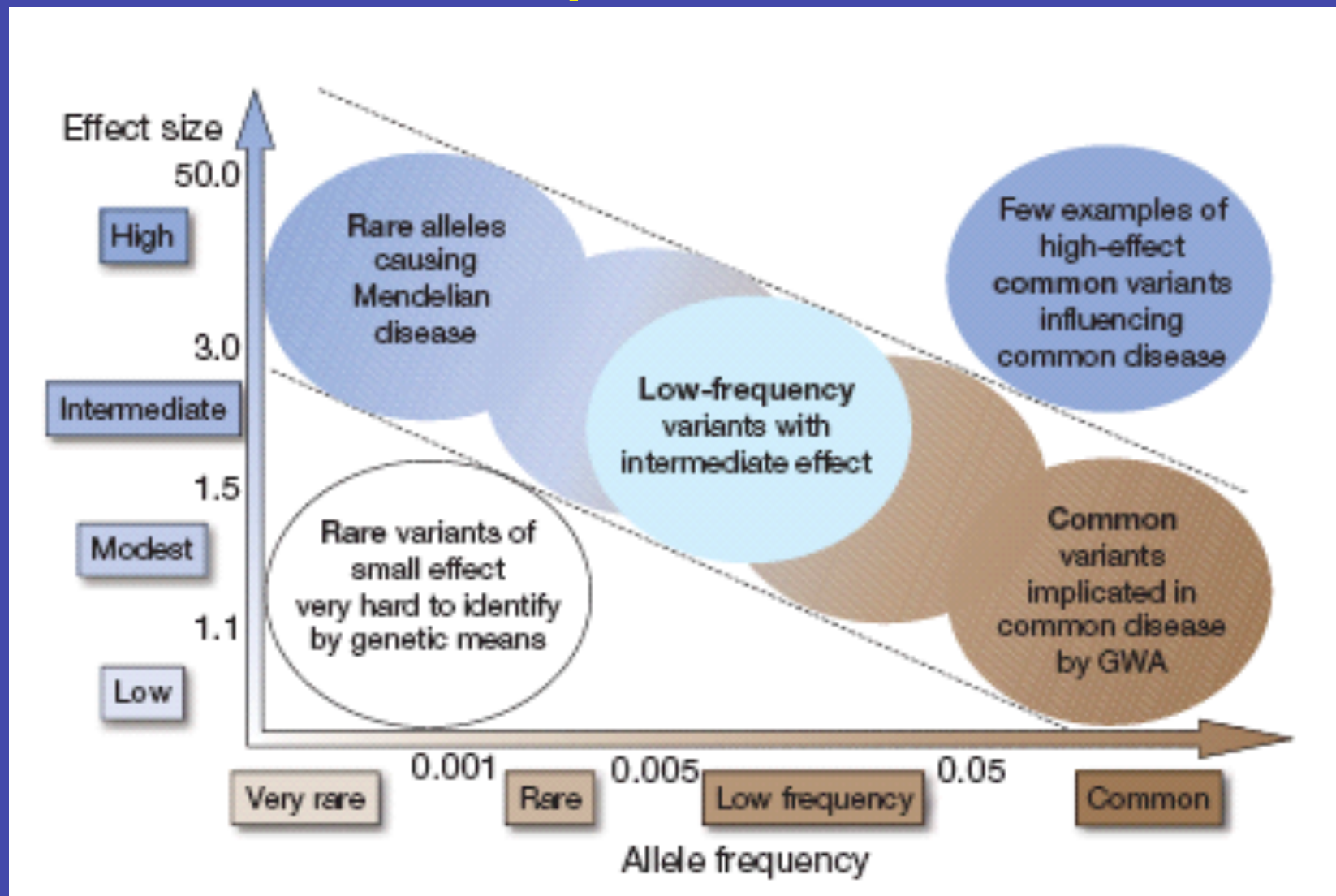
l'architettura di tratti genetici complessi > varianti di regioni non codificanti con ruolo nell'eziologia. Gli studi di linkage hanno stabilito: malattie complesse non determinate da pochi alleli rari con effetto forte. (selezione negativa degli alleli ad alta penetranza per riduzione di fitness).

GWAS hanno mostrato che gli effetti non possono essere spiegati da un numero limitato di varianti con effetto limitato (vedi figura sulla diagonale)

- determinare effetto quantitativo per l'associazione delle varianti a bassa frequenza isolate con GWAS

- varianti a bassa frequenza di effetti intermedi possono contribuire a spiegare la mancanza di fattori ereditari mancanti che dovrebbero emergere dalle grandi meta-analisi dei dati di GWAS

allelic freq vs. effect size



rapporto tra n.varianti e penetranza

progetto dei 1000 genomi

I progetti come quello dei 1000 sequenziamenti di genomi dovrebbero chiarire le informazioni sugli alleli a più basse frequenze e semplificare gli studi di WGA ad alto spettro di densità evitando la necessità del risequenziamento

le nuove tecniche di WGA e risequenziamento diranno se saranno ancora necessarie entrambe

esiste già una nuova generazione di tecniche per sequenziamento, il metodo Maxam e Gilbert sembra ormai un metodo della preistoria.

miglioramenti e prospettive

i WGA finquì non sono stati esaustivi

- hanno trovato scarsa % delle componenti genetiche delle eziologie multifattoriali
- necessità di maggior sforzi con ricerche più vaste
- a breve per trovare le mancanti componenti genetiche: aumentare le analisi sui fenotipi estremi, con famiglie informative disponibili (ricontattabili), espansione dei campioni per aumentare la copertura (più fitta), aumento delle varianti strutturali, meta analisi con soggetti di derivazione non-Europea
- creare nuova informatizzazione con il contributo funzionale delle varianti di: sequenza, struttura primaria e di cromatina, sensibilità ambientale, influenza del LD e degli aplotipi
- associazione dei loci con fenotipi multipli e identificazione dei tratti a maggior rilevanza per vari tipi di patologie (acqua calda).

nuove strategie:

Detection of associations with low frequency and rare variants will be facilitated by the comprehensive catalogue of variants with $MAF < 1\%$ being generated by the 1,000 Genomes Project (<http://www.1000genomes.org/page.php>), which will also identify many variants at lower allele frequencies. The pilot effort of that program has already identified more than 11 million new SNPs in initially low-depth coverage of 172 individuals ref.⁴⁴.

abbassamento dei costi e miglioramento delle tecniche,
passaggi diversi della risoluzione dei metodi:

WGA e resequencing

nuove strategie

associazione tra bassa frequenza e varianti rare sarà facilitata dalla catalogazione delle varianti con MAF (minor allele freq.) $< > 1\%$ ottenute col progetto 1000 genomi

<http://www.1000genomes.org/page.php>

che identificherà anche molti alleli a frequenze più basse.
L'effetto pilota del progetto ha già identificato 11×10^6 SNPs con una copertura su 172 soggetti

successivi abbassamento dei costi e rapidità di analisi stabiliranno il metodo che verrà adottato per screening routinari WGA-Reseq.

obiettivi

manca di completezza sulla informazione delle componenti genetiche spiegata con la numerosità dei tratti a bassa penetranza e poca informazione dell'architettura genetica dei tratti coinvolti

- determinazione delle funzioni dei tratti identificati
- identificazione delle varianti con le funzioni dei tratti mancanti per completare i fattori ereditari del fenotipo multifattoriale
- caccia dei fattori di rischio a piccolo effetto

risultati attesi

- identificazione delle centinaia di varianti di rischio a piccolo effetto
- identificazione delle piccole proporzioni di popolazione ad alto rischio genetico
- studio di nuove strategie di prevenzione mirata
- le varianti che partecipano all'insorgenza di più patologie determinate tramite WGA e reseq. di numeri ampi di pazienti e con i confronti dei diversi screening mostreranno i soggetti a rischio almeno per una patologia associata
- questo svilupperà la diagnostica e terapie sicure individualizzate benchè molto dibattute
- essendo piccola la descrizione della componente genetica dimostrabile delle malattie comuni le prossime ricerche sulle parti mancanti potranno dare molte informazioni sulle interazioni biologiche con componente genetica

ma dopo gli screening cosa viene?

pareva la panacea

- non si può tornare al riduttivismo
- analisi delle interazioni (deboli o forti)
- più che altro multiple, pleiotropiche ed epistatiche

analisi pleiotropica su “kindreds” = parenti

correlazioni dei livelli LDL TG, HDL-C

cardiovascular diseases CVD genetic risk (quantitativo)

low density lipoprotein LDL particle size	}	sottointendono rischio genetico
plasma triglyceride TG		
high density lipoprotein cholesterol HDL-C		

familial combined hyperlipidemia FCHL in coronary heart dis.CHD = variabile
familial monogenic hypertriglyceremia = obbligatoria

measures : LDL, plasma TG, HDL-C

Lipid Res. Clinics Family Study: atherosclerotic ph. CVD
association small dense LDL, low plasma level HDL-C, high
plasma TG, elevated apo-B levels

*max - likelihood multivariate quant.genet.

values obs.:	Relatives		Spouses		Total	
	Men (n 255)	Women (n 298)	Men (n 83)	Women (n 144)	Men (n 338)	Women (n 442)
age y	46.36±15.1	45.96±14.9	54.16±16.3	57.46±13.6	48.26±15.7	49.76±15.5
LDL size	262.46±9.2	267.56±8.3	264.56±9.2	268.46±8.9	262.96±9.2	267.86±8.5
TG mg/dl	185.56±136.0	144.36±91.2	158.56±130.8	145.66±89.0	178.86±135.1	144.76±90.4
HDL-C “	40.96±11.3	52.56±15.6	41.46±11.4	52.36±14.2	41.46±11.3	52.86±15.2

*Phenotype Pairs	rG6 ± SE	rE6 ± SE	rP
LDL size–TG	-0.876±06	-0.5360±05	-0.66
LDL size–HDL-C	+0.656±09	+0.486±06	+0.54
TG–HDL-C	-0.546±09	-0.536±07	-0.53
	additive genetic	environmental	phenotypic corr.

LDL low density lipoprotein

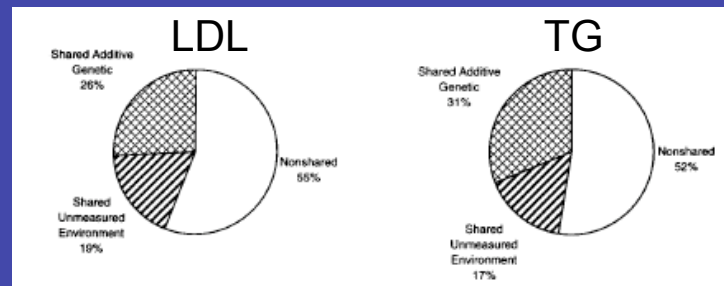
TG triglycerids

HDL-C high density lipoprot. cholesterol

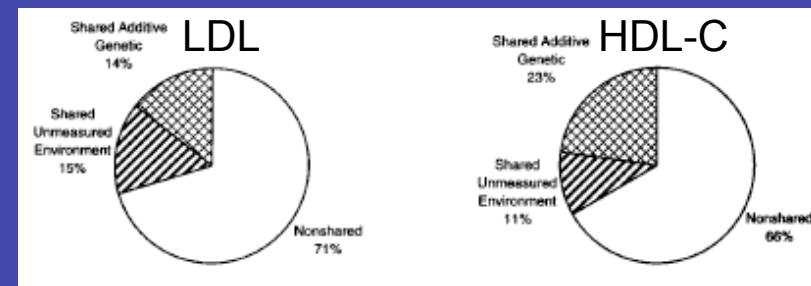
comparison of female monozygotic and dizygotic twins

lipid factor linked to: - lipoprotein lipase gene
- hormone sensitive lipase gene

common genetic evidences for the risk factor of CVD



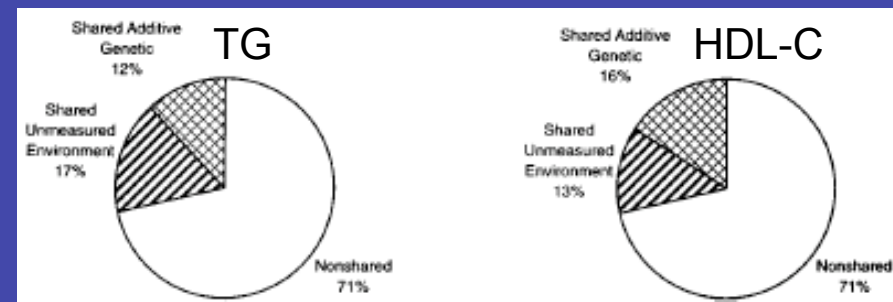
variance analysis



quantitative genet. analysis

quantitative genet. analysis

TG trigliceridi, LDL low dens. lipoprot.,
HDL-C high dens. lipoprot. cholester.



altogether

Together these findings suggest that each of the 3 lipid and lipoprotein measures reflect a common underlying process that is controlled, at least in part, by shared genes and provide strong support for the existence of genes with pleiotropic effects influencing covariation in plasma lipids and lipoproteins.

The use of multivariate traits is an important alternative to traditional approaches of estimating independent effects when the traits have a common etiologic pathway. This may be particularly true when considering genetically related risk factors.

both HDL-C and plasma TG included as covariates in analyses of LDL size account for shared additive genetic and random environmental contributions to the variance in LDL size.

overall

Evidence for linkage with the residual trait would represent a gene unique to that trait. Importantly, the pleiotropic effects identified in this study can thus be used to effectively increase the likelihood of detecting quantitative trait loci.

these results suggest that correlations between LDL size, plasma TG, and HDL-C are strongly influenced by shared genes and thus may be jointly involved in genetic susceptibility to CVD. Localizing and identifying these gene(s) may lead to a better understanding of the role of LDL size, plasma TG, and HDL-C in susceptibility to CVD in these high-risk families.

CVD cardio vascular disease

K.L.Edwards et al. Pleiotropic genetic effects on LDL size, Plasma triglyceride and HDL Cholesterol in families; Arterioscl. Tromb. and Vascular Biol. 1999, vol 19, 2456-2464.

epistatic effect

Methodology article Open Access Pub Med on line journal

Contribution of genetic effects to genetic variance components with epistasis and linkage disequilibrium

Tao Wang*1 and Zhao-Bang Zeng2

BMC (bio med central) Genetics 2009, 10:52 doi:10.1186/1471-2156-10-52