The Application of Molecular Epidemiology

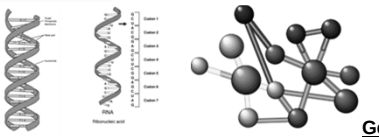
from virus transmission to the Human Behavioral Genetics Study

Yu-Ching Lan

yclan@ucla.edu
Visiting Scholar, Community Health Sciences, School of Public Health, UCLA
Assistant Professor, Department of Health Risk Management, School of Public Health, China Medical University, TW

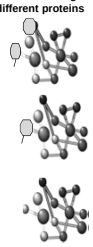
A biomaker exist in every living things

cgtcaatttagccctgactgtgtgtgaagaa
tacgcagttaaatcgggactgacacacacttctt



Gene Diversity

DNA mutation generate different proteins



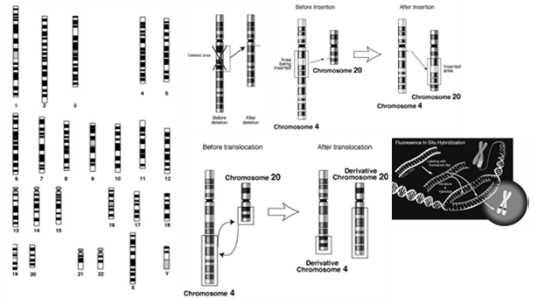
Second base

First base	U	C	A	G	Third base
U	UUU Phenylalanine F; UUC Phenylalanine F	UCU Serine S; UCC Serine S; UCA Serine S; UCG Serine S	UAU Tyrosine Y; UAC Tyrosine Y	UGU Cysteine C; UGC Cysteine C	U
	UUA Leucine L; UUG Leucine L		UAA Stop codon; UAG Stop codon	UGA Stop codon; UGG Tryptophan W	A; G
C	CUU Leucine L; CUC Leucine L; CUA Leucine L; CUG Leucine L	CCU Proline P; CCC Proline P; CCA Proline P; CCG Proline P	CAU Histidine H; CAC Histidine H	CGU Arginine R; CGC Arginine R; CGA Arginine R; CGG Arginine R	U; C; A; G
			CAA Glutamine Q; CAG Glutamine Q		
A	AUU Isoleucine I; AUC Isoleucine I; AUA Isoleucine I	ACU Threonine T; ACC Threonine T; ACA Threonine T; ACG Threonine T	AAU Asparagine N; AAC Asparagine N	AGU Serine S; AGC Serine S	U
	AUG Methionine start codon M		AAA Lysine K; AAG Lysine K	AGA Arginine R; AGG Arginine R	A; G
G	GUU Valine V; GUC Valine V; GUA Valine V; GUG Valine V	GCU Alanine A; GCC Alanine A; GCA Alanine A; GCG Alanine A	GAU Aspartic acid D; GAC Aspartic acid D	GGU Glycine G; GGC Glycine G; GGA Glycine G; GGG Glycine G	U; C; A; G
			GAA Glutamic acid E; GAG Glutamic acid E		

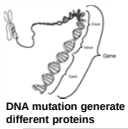
Gene Diversity I

different kind of gene variation

Cytogenetic map



4



Gene Diversity II

DNA mutation generate different proteins

point mutation

atgggtcaatttagccctgactgtgtgt

atgggtcaatttagccctgactgtgtgt

atgggtcaatttagccctgactgtgtgt

atgggtcaatttagccctgactgtgtgg

atgtgtcaatttagccctgactgtgtgt

atgtgtcaatttagccctgactgtgtgt

atgtgtcaatttagccctgactgtgttt

Can be selected for particular environments
(sickle hemoglobin cell - malaria)

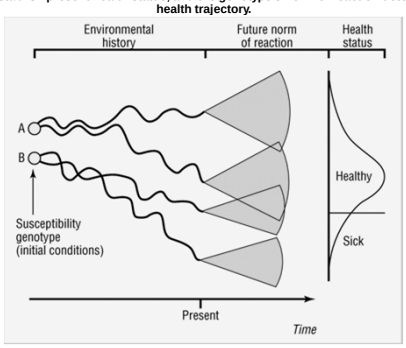
5

Effects of DNA Variants

- Silent
- Increase or decrease of gene product (protein)
- Change in how gene is turned on/off
- Change in protein function
- Change in Phenotype (such like behavior)

6

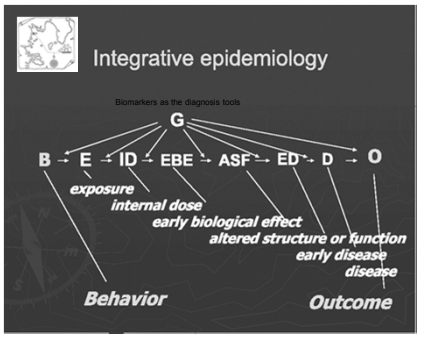
In complex disease a person's susceptibility genotype (such as A or B) and environmental history combine to establish present health status, and the genotype's norm of reaction determines future health trajectory.



Kaprio J BMJ 2000;320:1257-1259

BMJ

Integrative epidemiology is simply the familial molecular epidemiology paradigm with the 'wings', behavior, and outcome added.



Caporaso N E Cancer Epidemiol Biomarkers Prev 2007;16:365-366

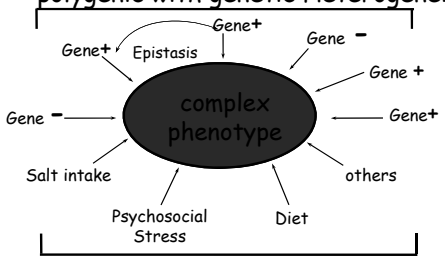
Cancer Epidemiology, Biomarkers & Prevention

- What molecular or genetic epidemiology care about?

Phenotype=genotype+environment+genotype*environment

"Heterogeneity " of complex diseases

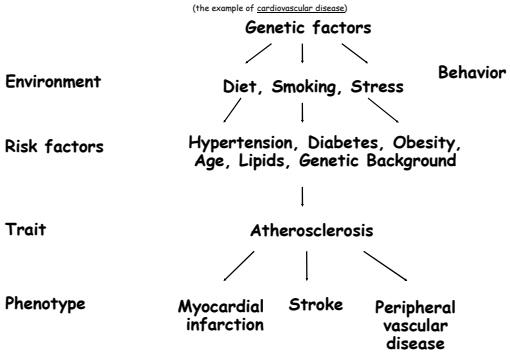
"polygenic with genetic Heterogeneity"



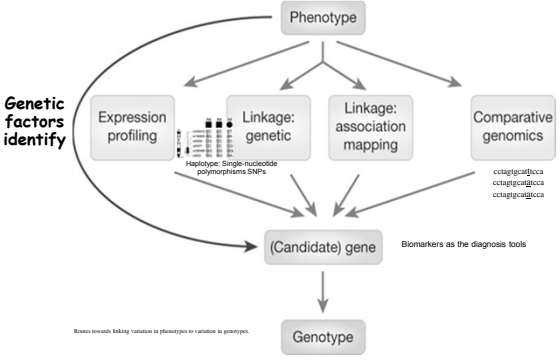
"Environmental factors"

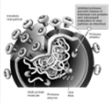
Gene-environment interactions

(the example of cardiovascular disease)



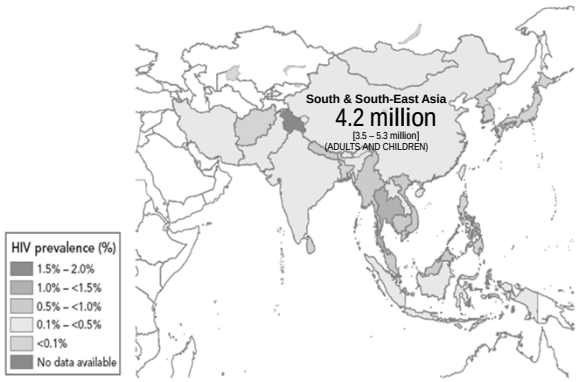
The strategies for searching the correlation between phenotype and genotype



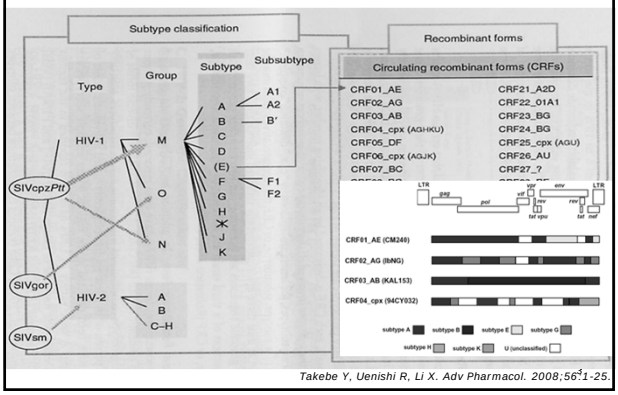


Example:
Virus genetic as biomarker
Comparative genomic study

Molecular Epidemiology of the Human Immunodeficiency Virus type 1 Infections in Taiwan



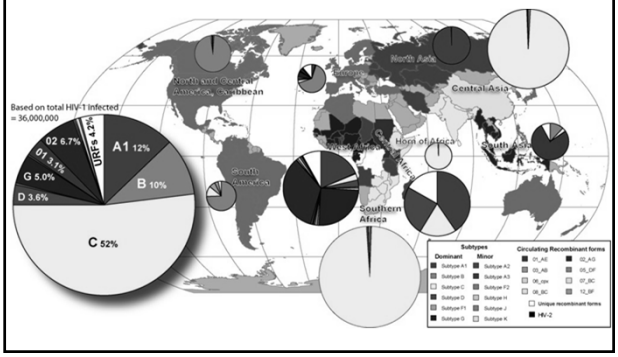
Classification of HIV



Takebe Y, Uenishi R, Li X. Adv Pharmacol. 2008;56:1-25.

HIV-1 subtype and recombinant prevalence in the world

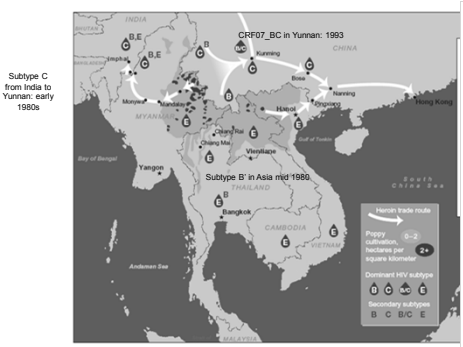
Subtype C is dominating the epidemic



Strategy for Tracing HIV Pandemic

- Epidemiological Data
- Medical Record Data
- Data from Genomic Diversity of the virus

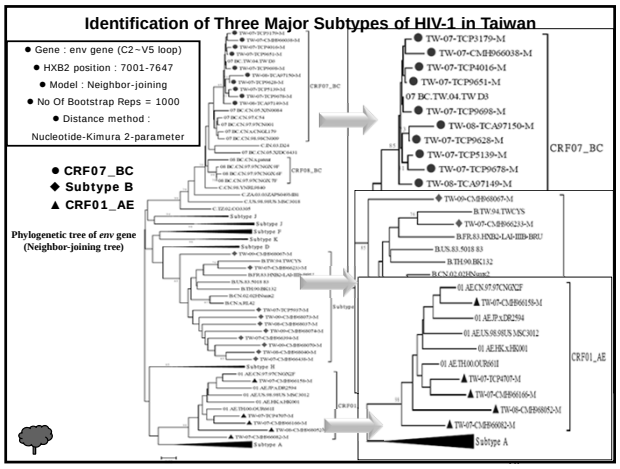
Infection pattern of CRF07_BC



Science. 2004 Jun 25;

Research Findings

- Subtype distribution after 2006
- Transmission origin and movement through Taiwan
- Characteristics among risk groups vis-à-vis HIV subtypes



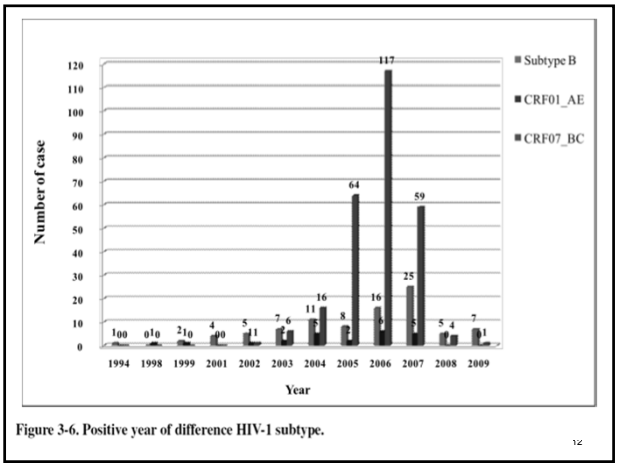


Figure 3-6. Positive year of difference HIV-1 subtype.

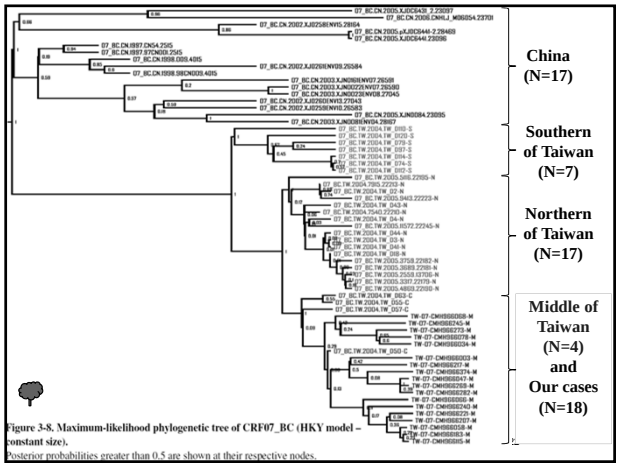


Figure 3-8. Maximum-likelihood phylogenetic tree of CRF07_BC (HKY model – constant size).
Posterior probabilities greater than 0.5 are shown at their respective nodes.

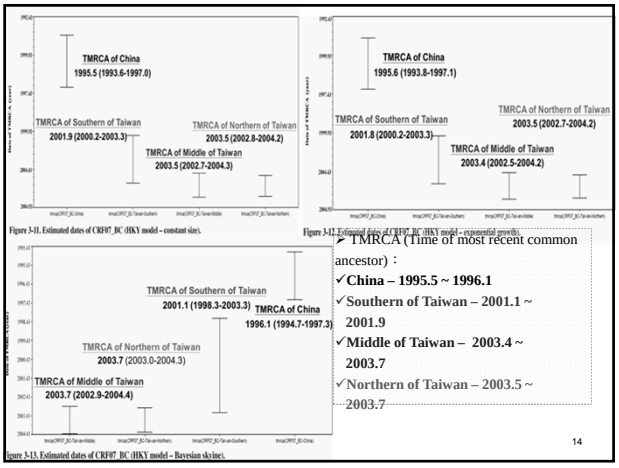


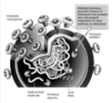
Figure 3-11. Estimated dates of CRF07_BC (HKY model – constant size).
Figure 3-12. Estimated dates of CRF07_BC (HKY model – exponential growth).
Figure 3-13. Estimated dates of CRF07_BC (HKY model – Bayesian skyline).

- TMRCA (Time of most recent common ancestor) :
  - ✓**China – 1995.5 ~ 1996.1**
  - ✓**Southern of Taiwan – 2001.1 ~ 2001.9**
  - ✓**Middle of Taiwan – 2003.4 ~ 2003.7**
  - ✓Northern of Taiwan – 2003.5 ~ 2003.7

Table 3-7. Comparison of HIV-1 subtype and demographic and risk behavior by questionnaires.

	Subtype B (N=39)	CRF01_AE (N=2)	CRF07_BC (N=73)	Total (N=114)	*p* value
	N (%)	N (%)	N (%)	N (%)	
Age (year)	(N=26)	(N=2)	(N=58)	(N=86)	0.007
Mean ± SD	30.7 ± 10.4	47.0 ± 17.1	37.2 ± 9.6	35.5 ± 10.4	
Education	(N=25)	(N=2)	(N=55)	(N=82)	<0.0001†
≤Junior high school	3 (12.0%)	1 (50.0%)	34 (61.8%)	38 (46.3%)	
≥Senior high school	22 (88.0%)	1 (50.0%)	21 (38.2%)	44 (53.7%)	
Marital Status	(N=25)	(N=2)	(N=56)	(N=83)	<0.0001†
Never been married	22 (88.0%)	0 (0.0%)	29 (51.8%)	51 (61.4%)	
Married/ Live together	2 (8.0%)	2 (100.0%)	7 (13.3%)	11 (13.3%)	
Divorce/ Separated/ Widow	1 (4.0%)	0 (0.0%)	20 (25.3%)	21 (25.3%)	
Sexual orientation	(N=22)	(N=2)	(N=49)	(N=73)	<0.0001†
Heterosexual	7 (31.8%)	2 (100.0%)	47 (95.9%)	56 (76.7%)	
Homosexual/Lesbian	13 (59.1%)	0 (0.0%)	2 (4.1%)	15 (20.5%)	
Bisexual	2 (9.1%)	0 (0.0%)	0 (0.0%)	2 (2.7%)	
Sexually Transmitted Disease (STD)	(N=21)	(N=2)	(N=53)	(N=76)	0.004†
Yes	8 (38.1%)	0 (0.0%)	4 (7.5%)	12 (15.8%)	
No	13 (61.9%)	2 (100.0%)	49 (92.5%)	64 (84.2%)	
Drug use	(N=25)	(N=2)	(N=55)	(N=82)	<0.0001†
Yes	4 (16.0%)	1 (50.0%)	49 (89.1%)	54 (65.9%)	
No	21 (84.0%)	1 (50.0%)	6 (10.9%)	28 (34.1%)	
Injection drug users	(N=25)	(N=1)	(N=55)	(N=82)	<0.0001†
Yes	0 (0.0%)	1 (100.0%)	50 (90.9%)	51 (63.0%)	
No	25 (100.0%)	0 (0.0%)	5 (9.1%)	30 (37.0%)	
Dilution sharing	(N=18)	(N=2)	(N=39)	(N=59)	<0.0001†
Yes	0 (0.0%)	1 (50.0%)	28 (71.8%)	29 (49.2%)	
No	18 (100.0%)	1 (50.0%)	11 (28.2%)	30 (50.8%)	
Container sharing	(N=18)	(N=2)	(N=39)	(N=59)	<0.0001†
Yes	0 (0.0%)	1 (50.0%)	23 (59.0%)	24 (40.7%)	
No	18 (100.0%)	1 (50.0%)	16 (41.0%)	35 (59.3%)	

† Fisher's exact test



Example2:
Virus genetic as biomarker
SNP study

Serological surveillance and IL-10 genetic variants on anti-HBs titers: Hepatitis B vaccination 20 years after neonatal immunization in Taiwan

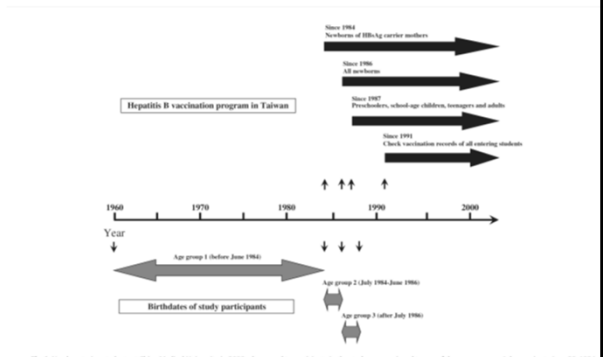


Fig. 1. Newly entering students at China Medical University in 2006 who agreed to participate in the study were assigned to one of three age groups: 1, born prior to June 30, 1984 (before Taiwan's hepatitis B vaccination program was initiated); 2, born between July 1, 1984 and June 30, 1986 (when the vaccination program only covered newborns of HBsAg carrier mothers); or 3, born after July 1, 1986 (when all newborns were covered).

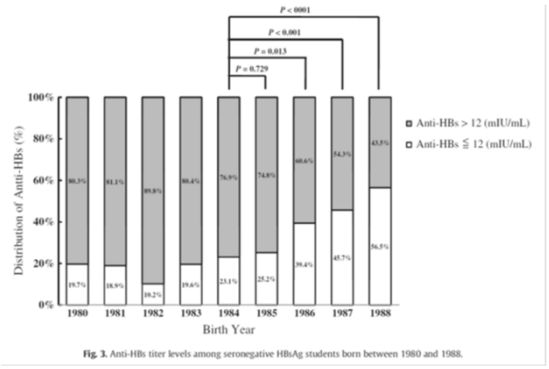


Fig. 3. Anti-HBs titer levels among seronegative HBsAg students born between 1980 and 1988.

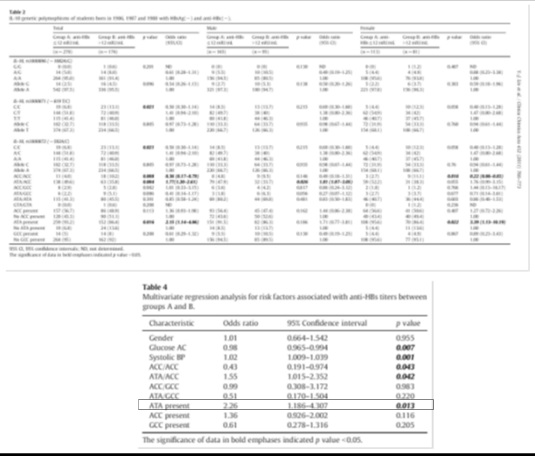


Table 4
Multivariate regression analysis for risk factors associated with anti-HBs titers between groups A and B.

Characteristic	Odds ratio	95% Confidence interval	*p* value
Gender	1.01	0.664–1.542	0.955
Glucose AC	0.98	0.965–0.994	**0.007**
Systolic BP	1.02	1.009–1.039	**0.001**
ACC/ACC	0.43	0.191–0.974	**0.043**
ATA/ACC	1.55	1.015–2.352	**0.042**
ACC/GCC	0.99	0.308–3.172	0.983
ATA/GCC	0.51	0.170–1.504	0.220
ATA present	2.26	1.186–4.307	**0.013**
ACC present	1.36	0.926–2.002	0.116
GCC present	0.61	0.278–1.316	0.205

The significance of data in bold emphases indicated *p* value <0.05.

How can we do in studying Genetics of Substance Abuse Behavior

The issues in substance abuse genetical study

- Understand the biologic factors of the illness
- Understand which proteins and pathways are contributing to particular mental illnesses
- Diagnosis
- Help to identify environmental factors (genes as co-variants)
- Risk Prevention
- Develop Better Treatments with Fewer Side Effects

Candidate gene search methods

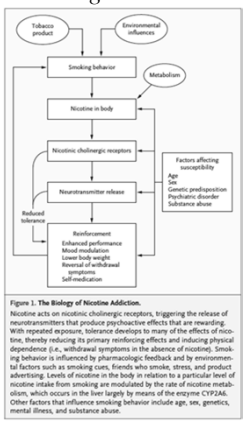
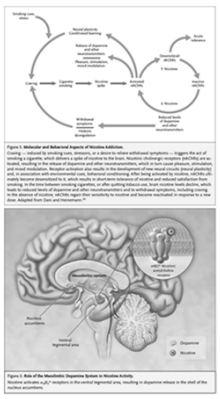


Figure 1. **The Biology of Nicotine Addiction.**
Nicotine acts on nicotinic cholinergic receptors, triggering the release of neurotransmitters that produce psychoactive effects that are rewarding. With repeated exposure, tolerance develops to many of the effects of nicotine, thereby reducing its primary reinforcing effects and inducing physical dependence (i.e., withdrawal symptoms in the absence of nicotine). Smoking behavior is influenced by pharmacologic feedback and by environmental factors such as smoking cues, friends who smoke, stress, and product advertising. Levels of nicotine in the body in relation to a particular level of nicotine intake from smoking are modulated by the rate of nicotine metabolism, which occurs in the liver largely by means of the enzyme CYP2A6. Other factors that influence smoking behavior include age, sex, genetics, mental illness, and substance abuse.

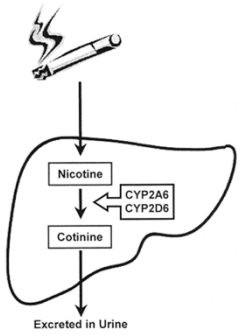


Nicotine Addiction

N Engl J Med 2010; 362:2295-2303June 17, 2010

Candidate gene search methods

Nicotine metabolism.



Batra V et al. Chest 2003;123:1730-1739

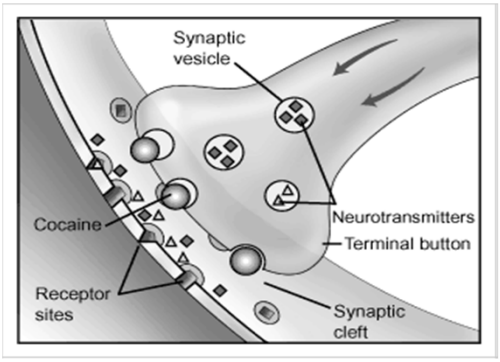
CHEST

Candidate gene search methods

Table 1—*Candiate Gene Association Studies in Smoking**

Association Studies	Sample Size	Current Smokers, %	Nonsmokers, %	Odds Ratio (95% CI)	p Value
Nicotine metabolism CYP2A6 polymorphism		Proportion of poor metabolizers			
Pianezza et al[35] (1998)	425 (Canada)	12.3	19.6	0.57 (0.34–0.96)	< 0.05
Tyndale and Sellers[36] (2001)	400 (Canada)	NA	NA	0.52	< 0.01
London et al[38] (1999)†	460 (US)	22.1	24.2	0.88 (0.56–1.39)	NS
Nicotine metabolism CYP2D6 polymorphism		Proportion of poor metabolizers			
Turgeon et al[39] (1995)	258 (whites, Canada)	0	10	NA	NA
Cholerton et al[41] (1996)	204 (Europe)	50	45.2	1.21 (0.70–2.10)	NS
Saarikoski et al[43] (2000)	976 (Finland)	2.1	2.3	0.91 (0.30–2.74)	NS
DRD2 polymorphism		Subjects (%) with A1 allele			
Noble et al[29] (1994)	354 (whites, US)	45.6	28	1.84 (1.19–2.88)	< 0.01
Comings et al[54] (1996)	1026 (whites, US)	49	26	2.71 (2.05–3.58)	< 0.001
Singleton et al[56] (1998)	221 (whites, UK)	31	43	0.59 (0.34–1.04)	< 0.1
Wu et al[55] (2000)‡	222 (AA and HA)	68.1	63.2	1.24 (0.61–2.53)	NS
Bierut et al[57] (2000)§	954 (US)	39.5	34.7	1.23 (0.94–1.60)	NS
Dopamine transporter SLC6A3 polymorphism		Subjects (%) with SLC6A3-9 allele			
Lerman et al[58] (1999)‖	522 (mixed, US)	46.7	55.8	0.52 (0.32–0.82)	< 0.01
Sabol et al[59] (1999)‖	1107 (US)	42	43.5	NA	< 0.05
Jorm et al[60] (2000)¶	561 (Australia)	46.5	46.2	0.93 (0.66–1.32)	NS

*CI = confidence interval; US = United States; AA = African American; HA = Hispanic American; UK = United Kingdom; NA = not available; NS = not significant.
†Data from patients enrolled in a case-control study for lung cancer.
‡Control population from a lung cancer case-control study.
§Family-based association approach.
‖Volunteer sample.
¶Nonvolunteer sample.

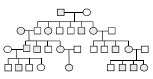


Study design

Family study

Genetic effect

Pedigree



Twin

environment effect

Adaptations Family migration Family

Individual study

- Molecular Genetic Studies
  - Association Studies
  - Linkage Studies
  - Whole Genome Scan
- Data collection
  - Epidemiological Data
  - Medical Record Data
  - Data from Genomic Diversity

Candidate genes using in addition research study

- DRD2
- DRD4
- COMT
- SERT
- BDNF
- *Ref : Candidate genes for cannabis use disorders: findings, challenges and directions. Addiction*, **104**, 518–532.