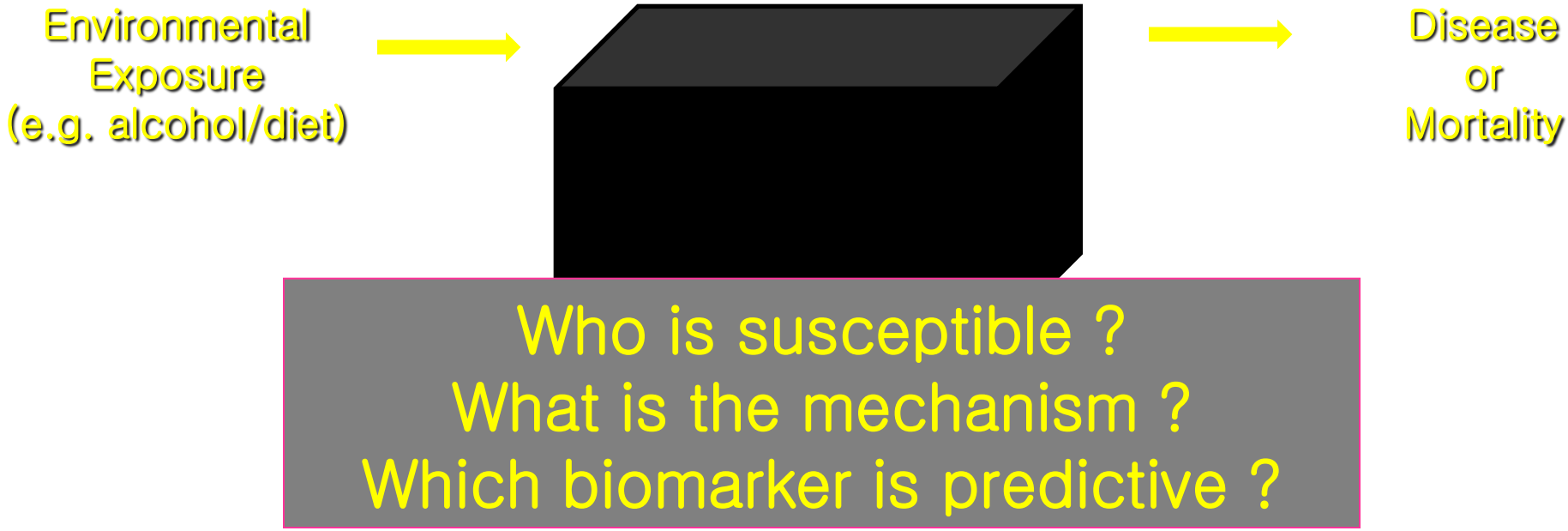
분자 - 유전체 역학연구

Molecular and Genome Epidemiologic Study

Dong-Hyun Kim

Hallym University College of Medicine

Evolving epidemiologic study …



What makes this box transparent ?

Origin of Molecular Epidemiology

- **The Impact of recent progress in molecular biology on epidemiology**
- **Based on general principles of epidemiology**
- **Utilizes molecular biology to define the distribution of disease and its etiologic**

Definition of Molecular Epidemiology

- **A science that deals with the contribution of potential genetic and environmental risk factors identified at the molecular level, to the etiology, distribution and control of disease in groups of populations**

- **Molecular biology << interface >> Epidemiology**

Epidemiology in the 21st Century: Population Impact of Human Genome Variation

- 19c ------ Population (Snow on Cholera ..)

20c ---------- Risk factor (Smoking ..)

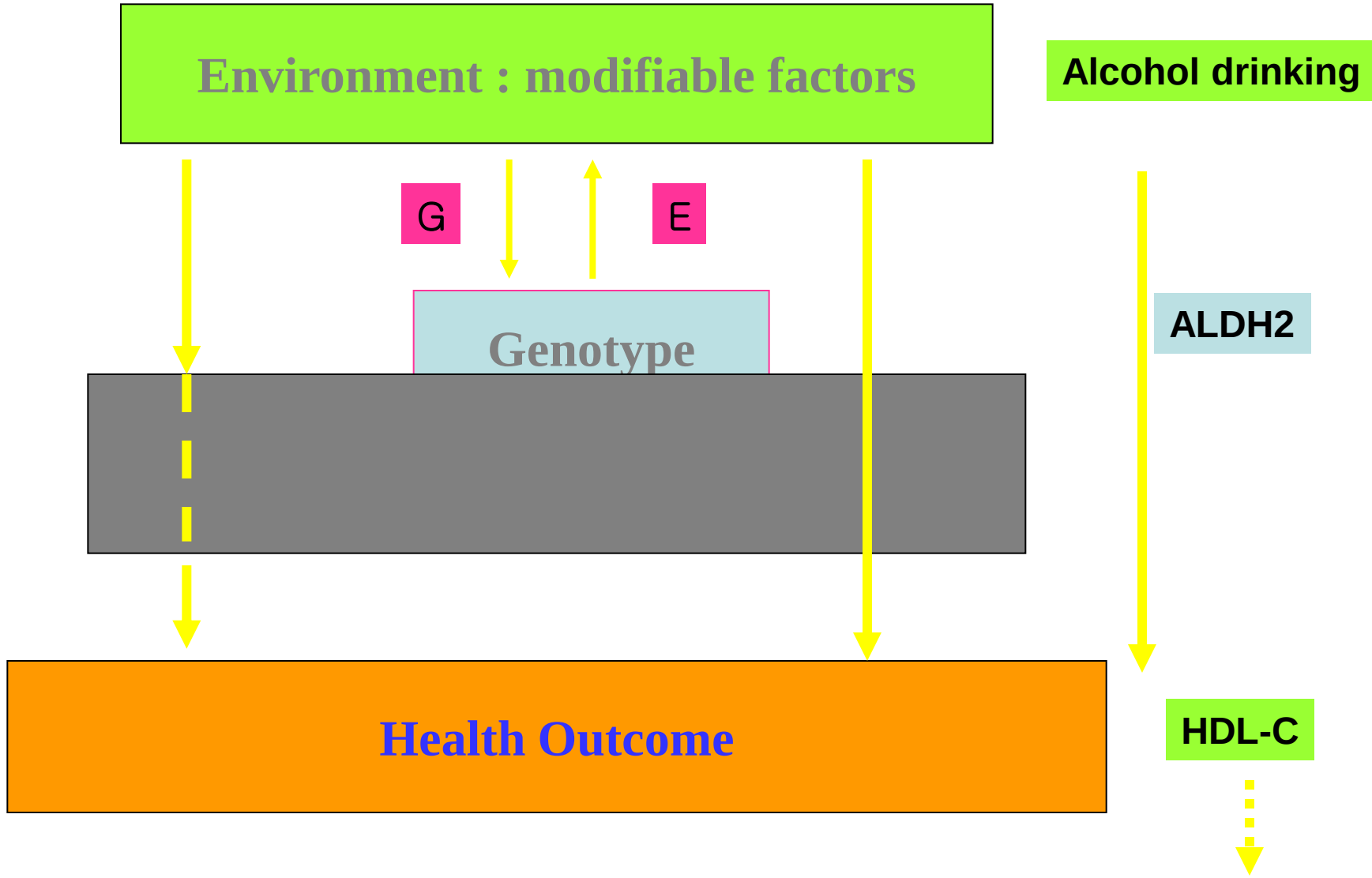
Success or Limit ?

21c ---------- Genetic variation

Tailored prevention ?

Gonomic Medicine

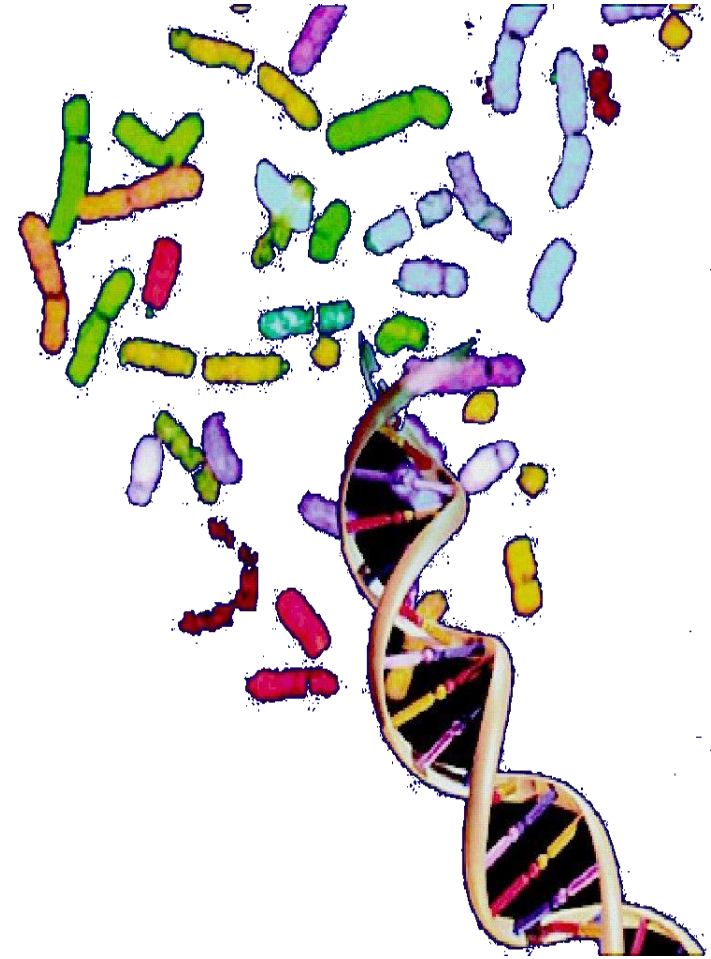
Epidemiologic Approaches to Human Disease In Transition



“Breakthrough of the Year”

- “For 2000, one word sums it up- GENOMES”

  – *Science* (2000)



"The Genome is Mapped: Now What?" (M.D. Lemonick, Time Magazine)

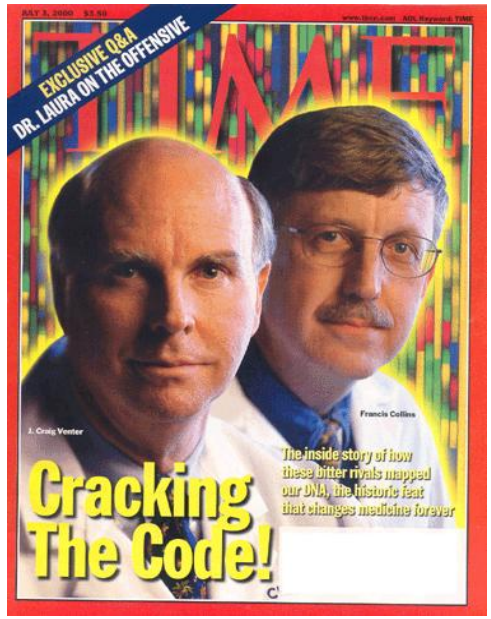
Gene Sequencing

Gene Discovery

Gene Characterization

Clinical Utility

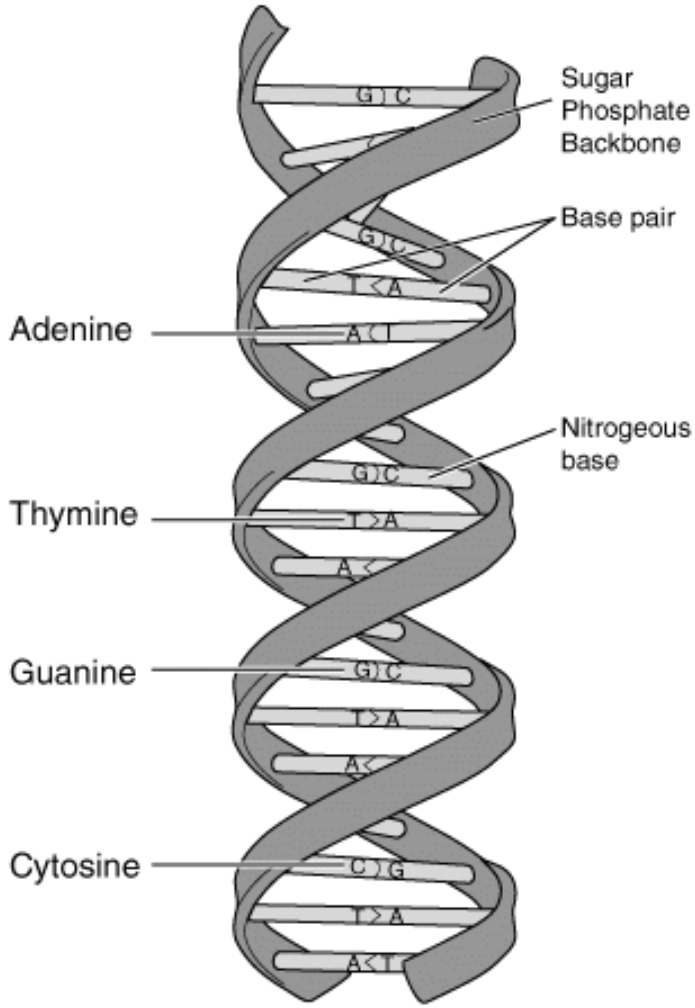
Public Health Utility



Emerging "Genomic" Methods in Epidemiology

- Susceptibility → DNA Polymorphisms
- Exposure → RNA Studies (Toxicogenomics)
- Outcomes → Disease characterization (Proteomics)

DNA Structure



Related Nomenclature

Nucleoside

Base + Sugar

Nucleotide

Base + Sugar + Phosphate

Nucleic acid

Linear sequence of Nucleotides

Human Genome

- The human genome consists of 3.2×10^9 basepairs of the nucleotides
- A rough draft of sequence of human genome has been completed
- The human genome has 30,000-40,000 genes encoding proteins
  - The gene is organized into

    1) DNA sequences involved in the regulation of gene transcription,

    2) DNA sequences encoding amino acids (exons)

    3) DNA sequences between exons (introns) whose function is unknown

DNA sequence differences/variants

- 99.9% of DNA sequences are identical among individuals
- 0.1% differences → on average, every 1000 bp
  - 1.68 million Single Nucleotide Polymorphisms (SNPs) were identified and made available to the scientific community (http://www.ncbi.nlm.nih.gov/SNP)
  - This number is expected to rise
- DNA sequence differences occurring with a frequency
  - < 1% are called mutation
  - >= 1% are called polymorphisms

Single-gene disease vs Complex disease

- Single gene disorder by causative gene, in which a mutation directly leads to the disease phenotype as in many Mendelian disorder (AD, AR, X-linked)
  - One-to-one correspondence between a given genotype and a given phenotype
  - High penetrance, but rare → high risk family syndrome

- Complex disease by susceptibility gene, that confers an increased risk of the disease phenotypes but which may not be sufficient in itself to result in the disease process
  - Polygenic disease : lack of one-to-one correspondence
  - Multifactorial : environmental factors interact with genetic predisposition
  - Low penetrance, but more common → higher population impact

Gene-environmental interaction

- Risk of a genotype may only be evident in exposed group
  - NAT1*10 and smoking vs bladder cancer

    - Crude analysis : NAT1*10 any vs WT Odds Ratio = 1.4 (1.0 - 2.1)

    - Considering Smoking :

Smoking	NAT1*10	OR
-	-	1 (ref)
-	+	1.1
+	-	2.4
+	+	6.0

Gene-environmental interaction

- Risk from environmental exposure may only be evident in genetically susceptible subgroups
  - Head injury and ApoE*e4

ApoE*e4	Head injury	OR (AD)
-	-	1 (ref)
-	+	1.0
+	-	1 (ref)
+	+	5.1

(Mayeux et at, 1995)

Gene-environmental interaction

- Identification of environmental “cofactors” that leads to clinical diseases in individuals with “susceptibility genotypes”

  → Future approach to the primary prevention of many diseases

(Khoury and Wagener, 1995)

Human Genome Epidemiology (HuGE)

- *by* Muin J. Khoury and Janice S. Dorman, CDC, USA

  (http://www.cdc.gov/genetics/hugenet/default.htm)

- An evolving field of inquiry that uses systematic applications of epidemiologic methods and approaches in population-based studies of the impact of human genetic variation on health and disease

- Intersection between molecular epidemiology and genetic epidemiology

Why population study ?

- **Better for unraveling the joint effects of environment and genes**
- **To assess the relative and attributable risk in the general population : public health genomics**
- **Greater statistical power to detect the genes with modest phenotypic effects on disease susceptibility**

(Risch et al, Nature 2000)

Why cohort study ?

- **Relatively free of selection bias**
  - **baseline cohort : representative pool of cases**
- **Nested case-control and case-cohort approach**
  - **Cost-efficient study design**
- **Permits analyses of different multiple outcome**
  - **Outcomes can be reliably compared, if case ascertainment is comparable**

Why cohort study ?

- Explore population stratification bias by treating the cohort as a “mini-population”
- Gene-environment interaction, using prospective exposure data
- The penetrance of the genes reliably estimated
  - accounting for risk factors

Perspectives of Human Genome Cohort

- **A very large population-based prospective cohort in the United States for the promise of the human genome project** **(Dr. Collins, Nature, 2004)**

- **What is urgently needed right now is a coordinated global initiatives to standardize and integrate data from the many cohort studies** **(Dr. Khoury)**

- **EPIC (Europe, ~500,000), MEC (USC, ~210,000), UKBiobank (UK, ~500,000), JMICC (Japan, starting), Lifegene (Singapore, starting) …. .**

Why Do We Need a Large Scale Genome Cohort Study in Korea ?

- To investigate the effects of genetic variants among Koreans and their interactions with environmental factors on the occurrence of several chronic disease and on the socially determined behavior,

- To understand the underlying mechanisms of the biological process of many diseases and to unravel complex pathways between life-style and socio-demographic factors and health,

- To set up the more tailored preventive strategy for target population with high risk of developing diseases, and

- To find pre-clinical biomarkers predicting the occurrence of various kinds of biological endpoints and social behavior

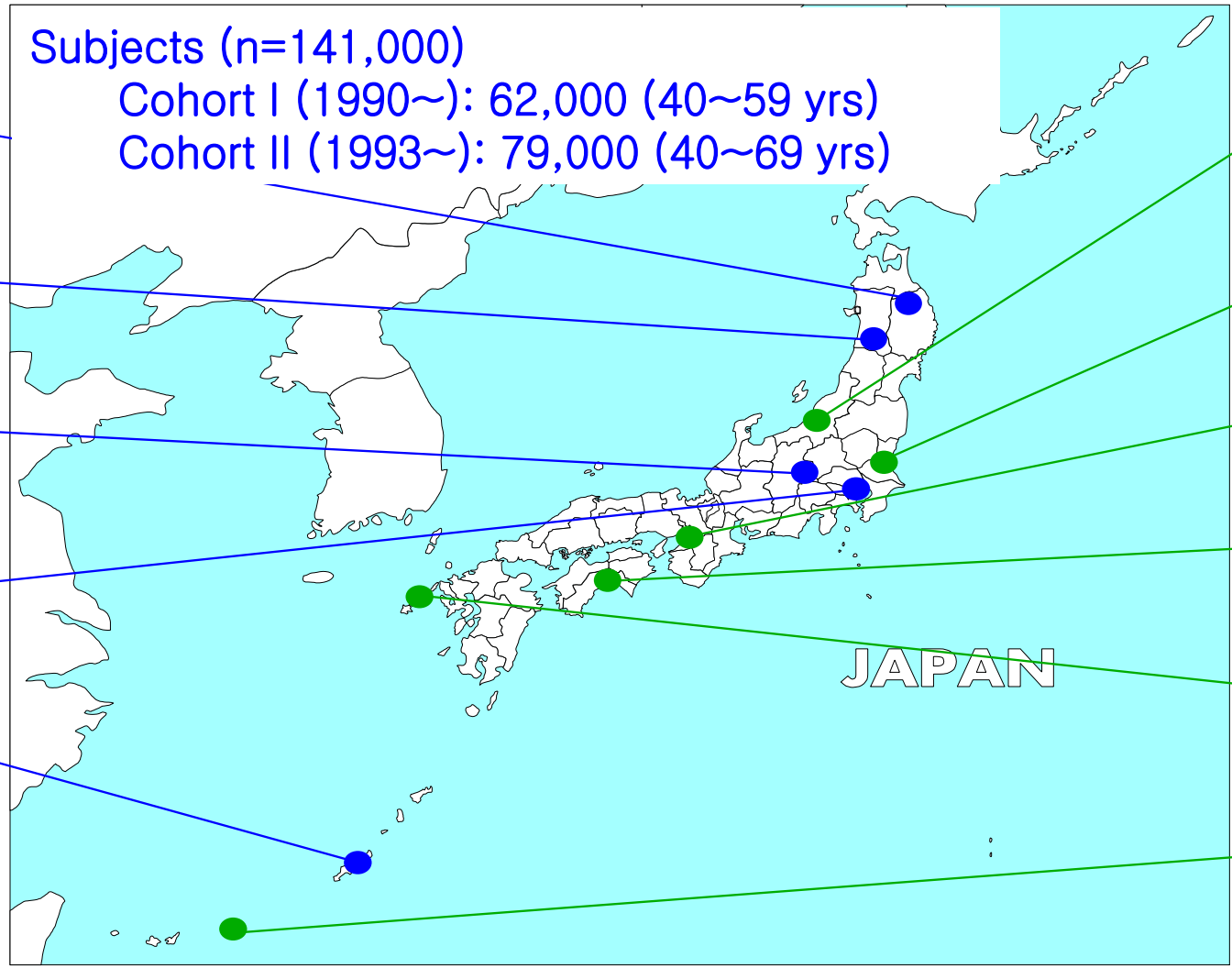
UK Biobank project (2002-9)
-a study on the genes, environment and health-

- 대규모 단일 코호트 구축 프로젝트
  - 50만명 이상의 대표성 있는 표본 선정 (45-69세)
    - GP: 500~600 선정
  - 자기기입식 설문
    - 사회경제적 상태, 인구학적 변수, 생활양식, 식이습관, 임신 관련요인, 질병 가족력
  - 임상검사 및 채혈
    - 혈압, 체격, FEV1
  - 추적조사 (암, 심장병, 당뇨, 치매)
    - NHS 기반
    - 통계청 자료 (사망률, 암 발생률) / 병의원 자료 (GP)

EPIC (1992~)
-European Prospective Investigation into Cancer and Nutrition-

- 목적: 식이요인, 생활양식을 포함안 환경적인 요인이 암과 만성질병발생에 미치는 영향 규명
- 연구참여국가: 10개국 (France, UK, Germany, Greece, Italy, Netherlands, Spain; Sweden Denmark, Norway(1995~))
- 연구대상: 520,000명 이상 (20세 이상)

JPHC (Japan Public Health Center-based prospective Study)



Cohort I

since 1990

- Ninohe (Iwate)
- Yokote (Akita)
- Saku (Nagano)
- Katsushika (Tokyo)
- Ishikawa (Okinawa)

Cohort II

since 1993

- Kashiwazaki (Niigata)
- Mito (Ibaraki)
- Suita (Osaka)
- Chuo-higashi (Kochi)
- Kamigoto (Nagasaki)
- Miyako (Okinawa)

* Public Health Center (Prefecture)

Korea Genome Epidemiologic Study

- Genome Epi cohort studies have been initiated since 2001

- Funded by the “Health Promotion Fund”, of the Korea Center for Disease Control

- Major types of study
  - Community-based study : Ansung(DM) and Ansan(Hypertension)
  - Large population study (LPS) in urban and rural area
  - Special group cohort, such as twin cohort, immigration study, elderly cohort, and growth/development cohort

- As of Dec. 2009, this on-going study has recruited 150,000 participants aged between 40 and 69 years-old from 40 centers scattered nationwide.