GBIO0002-1: Genetics and Bioinformatics
Homework 1b

DERYA KARAARSLAN

LUCA GUBERNATI

RAYAN KASSIR

Available online http://breast-cancer-research.com/content/9/6/114

Commentary

The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer

Christine B Ambrosone

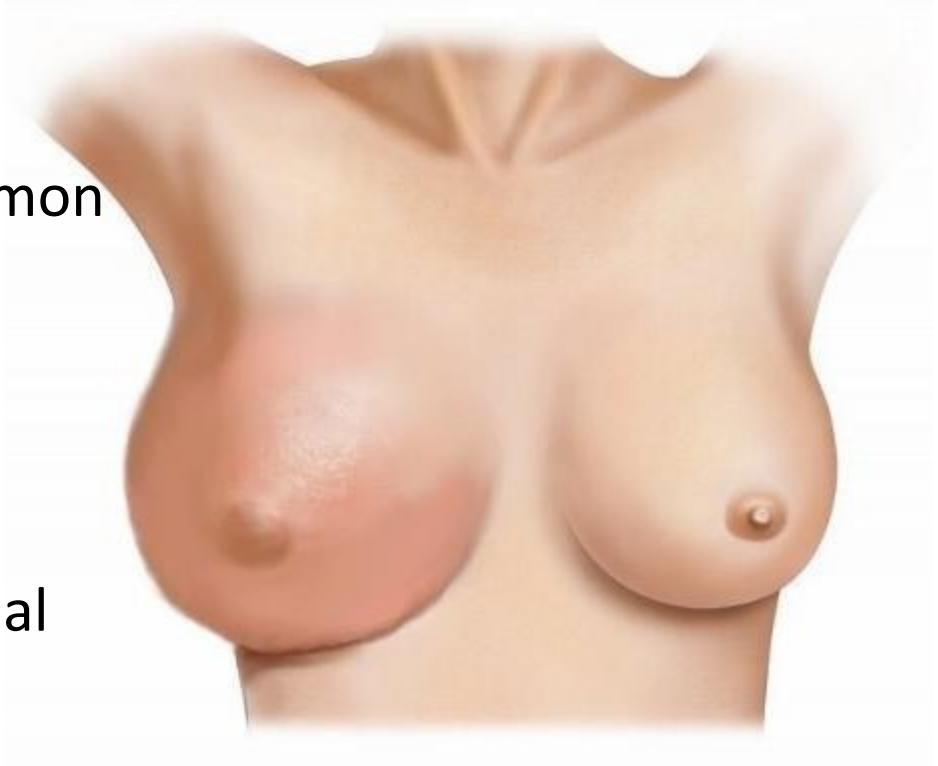
Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY 14263, USA

Corresponding author: Christine B Ambrosone, Christine.ambrosone@roswellpark.org

Published: 5 December 2007
This article is online at http://breast-cancer-research.com/content/9/6/114

Breast Cancer Research 2007, **9:**114 (doi:10.1186/bcr1787)

- Breast cancer is the most common form of malignancy affecting women worldwide.
- Appears to involve numerous genetic, endocrine, and external environmental factors.



- Genome-wide association = GWA
- Identification of cancer susceptibility alleles on populations,
- Certainly result in a great amount of new information and discoveries
- However the role of the environment, especially gene-environment interactions should also be taken into consideration whilst studying these results.

- 2xNormal woman = woman has relatives with breast cancer
- 30% of breast cancer cases because of genetic factors:
- *BRCA1*,
- *BRCA2*,
- *PTEN*,
- *p53*

%5 of global breast cancer

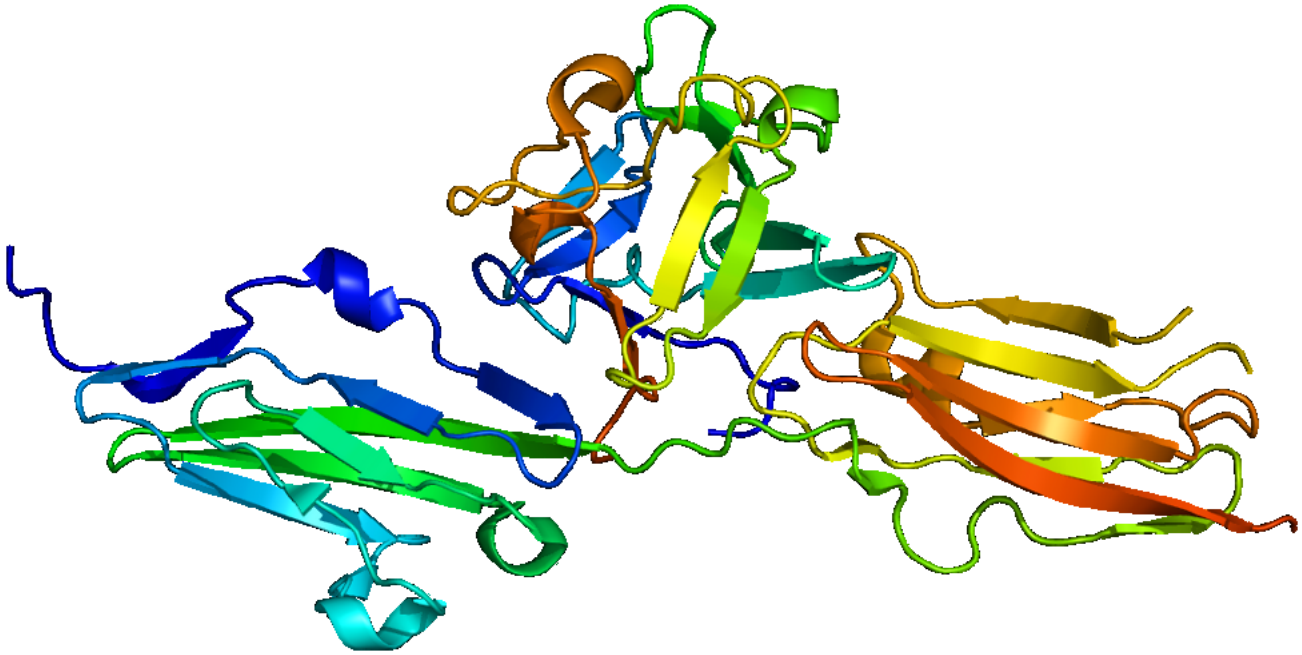
GWAS & DNA SEQUENCING → DIFFERENCES OF DNA SEQUENCE
SNPs

- Using a GWA study approach to breast cancer

➔ SNPs in fibroblast growth factor receptor 2

➔ regulating gene expression

- These results came similar in a group of 3 different studies based on three different types of populations



- Approach by comparing SNPs among cases with a strong family history of breast cancer with healthy control individuals to maximize the likelihood of identifying genes associated with inherited risk.

- The SNPs identified in fibroblast growth factor receptor 2 explained only 3.6% of familial risk, which would translate to much lower proportions in non familial breast cancer.

- The power of GWA studies in large sample sizes to identify gene variants that may increase risk of breast cancer, although these are not high-penetrance genes.

- The greater ramifications for these findings are that they identify pathways that have not been previously explored.

- **Limitations :**

- It is possible that some genetic variants that may be important susceptibility alleles are not covered by the SNPs that are genotyped.

- The probability that effects will only be noted in the context of exposures that induce expression of a gene or are associated with an increased or decreased risk as alcohol consumption, some dietary factors etc…

- Genetic variants and risks ➔obscure
- Complex systems of metabolism of numerous endogenous and exogenous compounds, that are possible factors and causes of breast cancer, are acknowledged when searching for the main effects of SNPs.
- Therefore, findings from GWAS studies for breast cancer will not result in identification of genes that explain a large portion of the breast cancer risk.
- The environmental effects on breast cancer risks by noting that rates of Japanese women with breast cancer increase within one or two generations after migrating to the US, to finally match rates of Caucasians.

- Studying gene-environment interactions is however difficult due to the complex mixtures of multiple factors to which humans are exposed.

- That being said, it is obvious that GWAS do not take into consideration the numerous factors as interactions between multiple genes and environmental exposures, in order to be capable of solving the mystery of breast cancer alleles.

- It is therefore advisable to proceed cautiously with the interpretation of studies designed to examine the effects of common SNPs on breast cancer risk, particularly in the absence of consideration of exposures

- GWAS will however, tend to be very useful in the cases of personalized therapeutics.

Hum Genet. Author manuscript; available in PMC 2013 October 01.

Published in final edited form as:
Hum Genet. 2012 October ; 131(10): 1525–1531. doi:10.1007/s00439-012-1209-8.

Study designs and methods post genome-wide association studies

Andreas Ziegler and
Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig–Holstein, Campus Lübeck, Maria-Goeppert-Str. 1, 23562 Lübeck, Germany

Yan V. Sun
Department of Epidemiology, Department of Biomedical Informatics, Emory University, Atlanta, GA, USA

Andreas Ziegler: ziegler@imbs.uni-luebeck.de; Yan V. Sun: yvsun@emory.edu

- GWAS :
  - case control studies×
  - Cohort studies

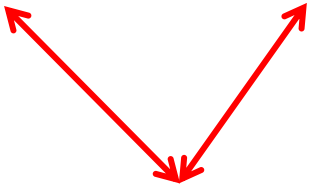
  extended families

  nuclear families

- This contrasts with the genome wide family based linkage studies that had much lower resolution with standard panels as it included only hundreds of microsatellite markers.

- Unexpectedly, many of the identified associations did not map to genes but to gene deserts.

- **coronary artery disease (CAD)** and **myocardial infarction (MI)**



9p21.3

- The success of GWAs using a large number of unrelated individuals :
  microarray technology & statistical methodology

The first set : composed of three articles (Almasy,Altmann and Wisjman)

- **Next Generation Sequencing (NGS)** which is the successor of GWAS.
- In the near future single strand whole genome sequences will become the final and standard DNA typing technology, allowing genetics to move from a technology-driven to a phenotype-driven scientific discipline.
- The post GWAS will turn to the phenotype, half of the phenotype-genotype connection. A wealth of data is available from high-throughput methodologies, such as transcriptomics, proteomics and metabolomics.
- Whole genome sequencing will most likely be the standard approach for genetic studies.
- Processes have yet to be standardized for whole genome sequencing as it was the case for GWAS in order to make it easier and help grow the data base internationally.

- **One of the major aims in sequencing studies is to identify genetic variations, such as SNPs and insertions/deletions, for further genetic analyses.**

- Steps and guide to SNP calling ➔ process begins with base calling, followed by quality control, mapping of the reads to the reference genome, and post alignment and post mapping and finally, the variant calling procedure itself.

- One important observation by Altmann is that **the final SNP calls are substantially affected by the choice of the specific software package used for SNP variant calling.** This proves the necessity to establish a reliable easy to use pipeline for variant calling.

- Clerget-Darpoux and Elston (2007) argued that the common disease-common variant GWAs approach fails to identify rare variants
- «multiple rare variants in a disease gene can play an important role in disease susceptibility.»

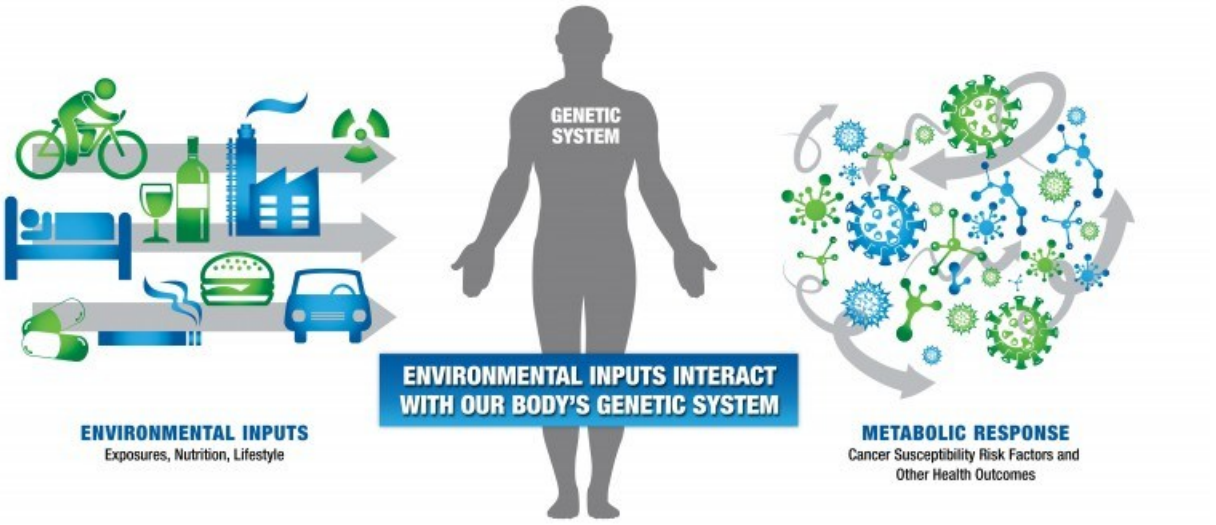
- Family information is relevant to refine the genetic model and for estimation of disease risk.
- But according to Wisjman, the large pedigree design is practical, efficient, and well suited for investigating rare variations as *it only requires a small fraction of the sample size needed to identify rare variants of interest not to mention the numerous and available statistical and computational tools that already exist.*
- One of it's most interesting advantages is the ready availability of samples of large pedigrees with large amounts of phenotypic data for high-throughput sequencing.

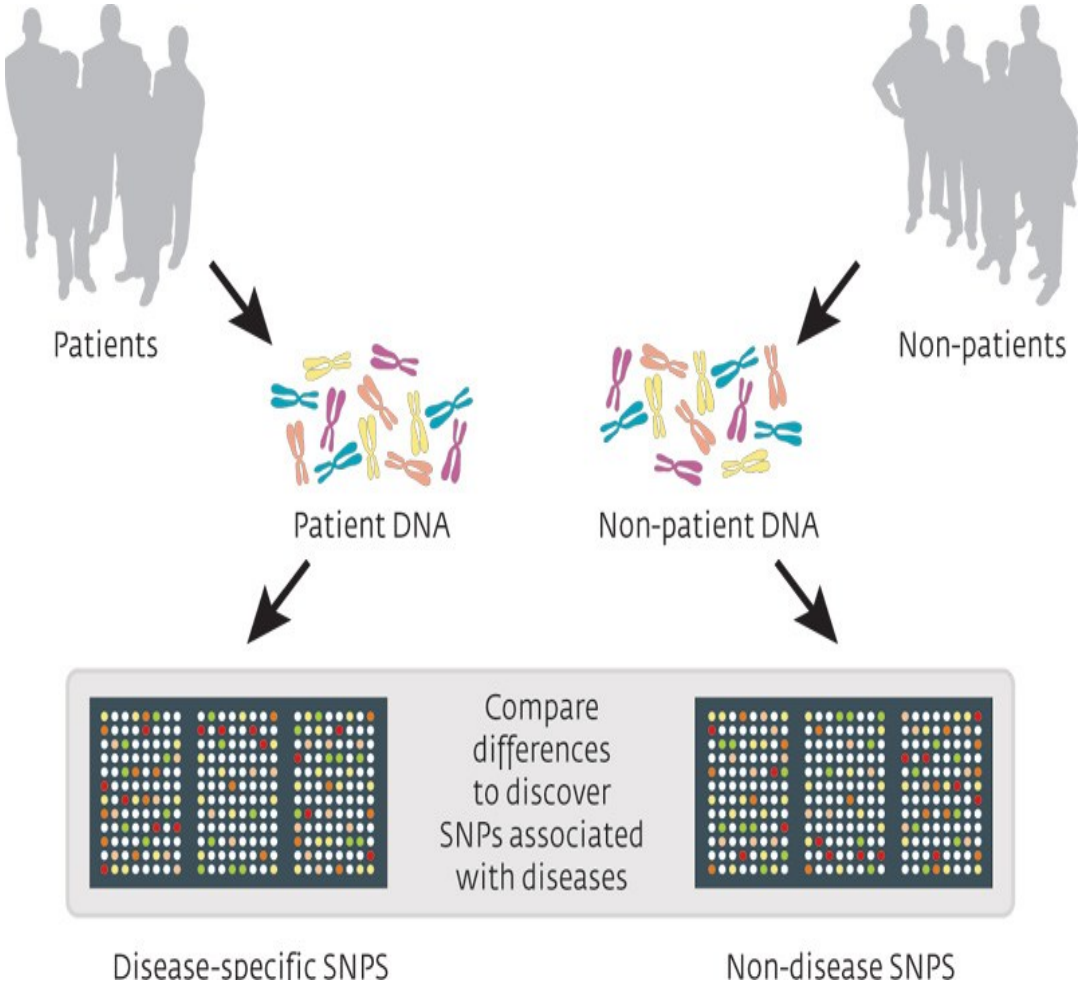
The second set is composed of articles on study design (Cortessis&Al) :

- The authors do not follow the broad definition of epigenetics that has recently been used «to describe any non-genetic mechanism influencing phenotype».
- They adopt more specific definitions, differentiating *epigenetic processes* that stably affect gene expression through mechanisms not involving the primary nucleotide sequence from epigenetic states, the configuration of chromatin and DNA marks utilized by these processes.
- With appropriate limitations to the scope of research questions and interpretation of molecular mechanisms, we can have an initial understanding of common diseases.
- However this faces many challenges : Measuring the exposure, The biospecimen itself etc..

- GWEI (genome-wide environmental interaction) studies are used to study the statistical effect of interactions between genetic markers and environmental factors on disease.

- Epigenetic studies primarily deal with the identification of differences in epigenetic profiles, which can be the molecular mediator of the gene-environment interaction, between two groups of individuals.





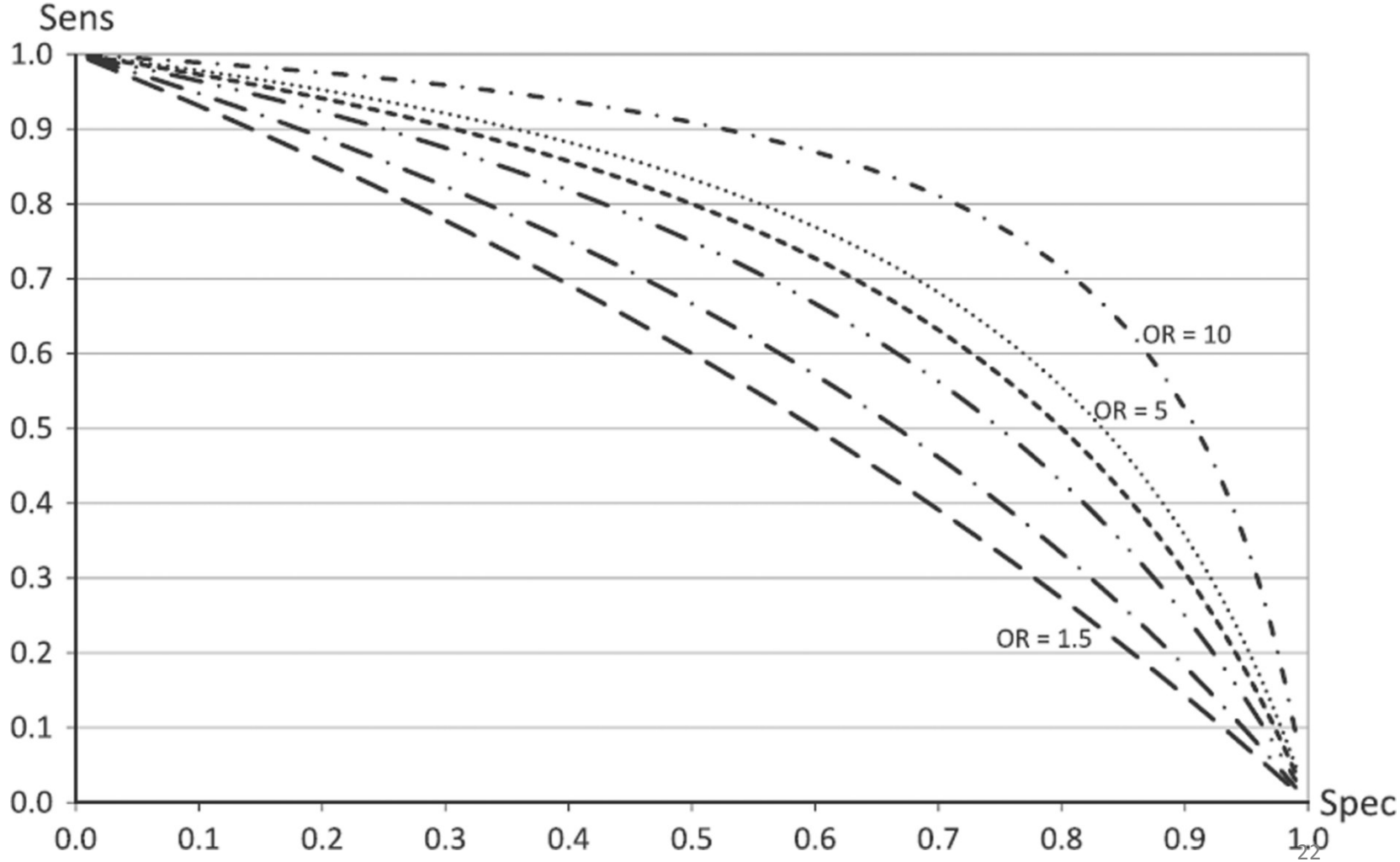
- Pharmacogenetic studies constitute a special case of gene-environment interaction studies, where interactions between genes and drugs are studied to improve treatment efficacy and/or to reduce adverse effects.
- The genotyping can be very targeted if the mechanism of action of the drug is known, and small-scale chips specific to these known mechanisms have been developed by several companies and academic groups.

The third set is composed of articles on analytical methods (Kruppa&Al) :

- Provides an overview on classification and risk estimation using machine learning methods.
- **Prognosis** and **diagnosis** using biomarkers, such as data from GWAs.
- SNPs, which may play important roles in disease etiology, can be poor classifiers, a phenomenon that can be observed from a relationship between the odds ratio (OR) and sensitivity (sens) and specificity (spec).

$$\mathrm{OR} = \frac{\mathrm{sens}}{1 - \mathrm{sens}} \times \frac{\mathrm{spec}}{1 - \mathrm{spec}}.$$

- Kruppa et al. argue that no single measure or classification is sufficient to judge the clinical validity of a “multimarker rule,”.
- Relationship between strength of association and classification accuracy. Sensitivity (sens) is plotted as a function of specificity (spec) for odds ratios (OR)



- Zaitlen and Kraft (2012) consider an entirely different application of GWA data, estimating the proportion of phenotypic variance explained by genetic factors .
- Heritability estimates using the traditional approach were biased and sometimes even overly optimistic.

- An alternative has recently been offered by Yang and Visscher in the context of GWAs (Yang et al. 2010), in which observed DNA variation from genotyping or sequencing is used for heritability estimation.

- However, most GWAs and most genetic epidemiological studies lack the familial component; therefore, causality needs to be considered by employing a different approach.

- Most GWAs only examine the association of a single SNP at a time and ignore the connectivity between the genes.
- Sun (2012) summarizes approaches to incorporating the pathway and network into the GWAs analysis.
  - On the one hand, pathway and network information may assist in identifying a set of functionally related genetic associations with a trait,
  - On the other hand, the statistical inference from the GWA data can also infer novel biological relationship among genes.

- NGS (next generation sequencing) is generating more data in a more cost-effective way than ever before.
- The deluge of data will only be totally when the corresponding issues in study design and methodology are well considered and properly developed.
- The success of GWAs demonstrates the tremendous value of efforts in design and method development, leading the way to the new necessary reaserches for the post GWAS era.

THANK YOU FOR YOUR ATTENTİON…