PHGEN I

Working Paper 1

Public Health
Public health is a social and political concept aimed at improving health, prolonging life and improving the quality of life among whole populations through health promotion, disease prevention and other forms of health intervention (1)

Public Health Trias
The Public Health Trias describes the three core-functions of Public Health: assessment, policy development, and assurance. These three core functions of Public Health have been transferred by McGinnis to the concept of Public Health Genomics (2):
- Assessment: to monitor the influence of genetic factors on population health
- Policy Development: to establish policies and guidelines that support sound and efficient application of genetic tools to improve populations health
- Assurance: to assure access, quality and appropriateness of genetic services

The trias is further described by the ten essential services of Public Health which have also been transferred to the concept of PHG by McGinnis

Empowerment
Empowerment is a process through which people gain greater control over decisions, and actions affecting their health (1)

Prevention
Prevention can be differentiated as primary prevention, secondary prevention, and tertiary prevention. Primary prevention is any activity by which an individual avoids the development of a disease (primary prevention), diagnosis a disease in an early stage or prevents its reoccurrence (secondary prevention), or avoids a disease's worsening and restores oneself to an optimal level of functioning (tertiary prevention).

Determinant of health
Any factor, whether event, characteristic, or definable entity, that brings about change in a health condition or other defined characteristic (e.g. Physical and social environmental factors, biological and genetic factors, factors relating to the health care system, lifestyle factors) (3)

Epidemiology
The study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems (3)

Human Genome Epidemiology
Human Genome Epidemiology (HuGE) aims to identify genome-based health information of public health relevance by analysing the impact of genomic variants on the risk of developing diseases or disabilities and to measure how other risk factors (e.g. lifestyle or environment) interact with genomic variants (4)

Validity, analytical

The analytical validity expresses the reliability and accuracy of the assay. In the context of genetic testing it is a measure of the ability of the assay to measure the genotype of interest. The evaluation process requires precise definition of the genotype or set of genetic variants that comprise the assay (5).

Validity, clinical

Conflates two separate entities (5):
  ▪ In the context of genetic testing the assessment of the link between genotype and disease (which in many cases is addressed through a systematic review of genetic association studies)
  ▪ The evaluation of test performance (by which we mean studies to determine characteristics such as sensitivity, specificity, PPV, NPV and likelihood ratios).

Sensitivity

Sensitivity is the proportion of truly diseased persons in the screened population who are identified as diseased by the screening test. Sensitivity is a measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test (3).

Specificity

Specificity is the proportion of truly non-diseased persons who are so identified by the screening test. It is a measure of the probability of correctly identifying a non-diseased person with a screening test (3).

Predictive Value

In screening and diagnostic tests, the probability that a person with a positive test is a true positive (i.e., does have the disease) is referred to as the "predictive value of a positive test. The predictive value of a negative test is the probability that a person with a negative test does not have the disease. The predictive value of a screening test is determined by the sensitivity and specificity of the test and by the prevalence of the condition for which the test is used (3).

Risk factor

An aspect of personal behaviour or life-style, an environmental exposure, or an inborn or inherited characteristic, that, on the basis of epidemiological evidence, is known to be associated with health-related condition(s) considered important to prevent (3).

Clinical utility

Clinical utility is a measure of the health care value of provided by a test or a technology (5).

Health Technology Assessment (HTA)

Health technology assessment (HTA) is defined by the European Network for health Technology Assessment (EUnetHTA) as multidisciplinary process that summarizes information about the medical, social, economic and ethical issues related to the use

of a health technology in a systematic, transparent, unbiased, robust manner. Its aim is to inform the formulation of safe, effective, health policies that are patient focused and seek to achieve best value. Despite its policy goals, HTA must always be firmly rooted in research and the scientific method (6).

Health Need Assessment (HNA)
HNA is a systematic method of reviewing the health issues facing a population, leading to agreed priorities and resource allocation that will improve health and reduce inequalities (7).

Health Impact Assessment (HIA)
HIA is defined in the Gothenburg consensus as combination of procedures, methods and tools by which a policy, a program or project may be judged as to its potential effects on the health of a population and the distribution of effects within a population (8).

Public Health Genomics (PHG)
Public Health Genomics is defined by the Bellagio Group as the responsible and effective translation of genome-based knowledge and technologies into public policy and health services for the benefit of population health (9).

References

1. World Health Organization (WHO). Health Promotion Glossary. Geneva: World Health Organization; 1998.
2. McGinnis J. The Public Health System. In: Hernandez L, editor. Implications of Genomics for Public Health - Workshop Summary. Washington D.C.: Institute of Medicine; 2005. p. 32-4.
3. Last J. A Dictionary of Epidemiology. Oxford University Press; 2001.
4. Khoury M.J., Little J., Burke W. Human Genome Epidemiology: scope and strategies. New York: Oxford University Press; 2004.
5. Burke W., Zimmern R. Moving beyond ACCE: An expanded framework for genetic test evaluation: PHGFoundation; 2007.
6. EUnetHTA. HTA. European Network of Health Technology Assessment (EUnetHTA); [cited 2007 November]; Available from: http://www.eunethta.net/
7. Health Development Agency. Health needs assessment - a practical guide. London: Health Development Agency; 2005 [cited 2007 November].
8. Gothenburg Consensus Paper. WHO European Centre for Health Policy; 1999 [cited November 2007]. Available from: http://www.euro.who.int/document/PAE/Gothenburgpaper.pdf
9. Burke W, Khoury MJ, Stewart A, Zimmern RL. The path from genome-based researchto population health: development of an international public health genomics network. Genet Med. 2006 Jul;8(7):451-8.

Working Paper 2

Genetic Diseases

Meaning diseases in which gene mutations are to a significant extent causally involved. However, genes are involved in all disease, as they are also involved in the physical and mental status we call health.[1] In conclusion, the term "genetic disease" has no meaning, because it includes any disease: the prefix "genetic" does not indicate a subset of the group referred to as "disease".

Inherited Disorders

Defined as a disease caused by inherited factors producing the disease in most persons with those genetic factor(s), irrespective of their other genes and their environment. Such genetic factors may be denoted "genetic defects". "Mutations" may be genetic defects, but most mutations are not. The term "mutation" indicates today a sudden and inherited change in DNA structure.

In order to distinguish rare "inherited disorders" from common diseases, we use instead the term "common-complex diseases" to refer to the disorders, such as cancer or heart disease, that are the focus of public health practice We will differentiate and use the following terms in the following way:
- *Inherited disorders* are Mendelian disorders with high penetrance. Their prevalence tends to be low.
- *Common-complex diseases* have wide prevalence. Genetic variations are or are expected to be identified that contribute to developing the disease

Genetic Information

Genetic information is information about a person's genetic makeup. However, this information might be derived from many sources, not only analysis of DNA or RNA. The source does not say much about the moral or social relevance of the property of the information gained. In other words, the handling of the information should be decided by the implications of the information, not by the source of the information.
Genetic information about inherited disorders, though, is of a different morally relevant quality, primarily because of its high degree of predictability and its relevance for other family members

Genetic Testing

Refers to any test/diagnosis that delivers genetic information; including conventional testing.

Gene Testing

Is a subgroup of genetic testing. Gene testing is the analysis of DNA or RNA.

---

[1] Cf. Smith, Kelly (2001) A Disease by any other Name: Musing on the Concept of a Genetic Disease. In: Medicine, Health Care and Philosophy 4: 19-30.

Working Paper 3

Genetic Screening
By contrast to genetic testing, the term genetic screening is used primarily for testing a large number of populations or sub-populations of mainly asymptomatic individuals who are tested if they might or might not be at increased genetic risk of a particular condition.

Surveillance
Surveillance is the regular, systematic collection, analysis and interpretation of information for use in tracking and describing changes in the occurrence of diseases over time.

PHGEN I

Working Paper 5

Genetic literacy
Means the general understanding of basic facts relating to the genome, probability of health information and patterns of inheritance.

Health literacy
Is the capacity to make better informed decisions with regard to ones own health – so that citizens can become better co-producers of their health. One aspect of genetic and health literacy is to know that *predictive knowledge* that is generated in the context of genome-based knowledge is *probabilistic knowledge*. As a consequence, persons have to be educated on how to deal with these uncertainties (although they are still dealing with uncertain probabilistic knowledge e.g. when knowing about high cholesterol or blood pressure).

Human Dignity
Inference derived from human dignity is understood as a negative right, a "hands off"-claim. Its normative power in the context of Public Health Genomics can be unfolded in this way. In more specific terms, it means in this context that persons do have a right to self determination especially in the context of their own health information and data. This includes the right to know or a right not to know about one's health predispositions and susceptibilities. Especially in this context,

For public health approaches it is certainly helpful to consider different principles as well, for example *health maximisation* (as a transformance of the beneficence-principle from individual to social ethics) and *proportionality*[2] as the conflict between collective interests/ rights and individual interests/rights is immanent to public health conflicts and have to be balanced proportionally.[3]

---

[2] Childress JF, Faden RR, Gaare RD, et al. (2002) Public Health Ethics: Mapping the Terrain. Journal of Law, Medicine & Ethics 30: 170-178. Thompson AK, Faith K, Gibson JL, Upshur REG (2006) Pandemic influenza preparedness: An ethical framework to guide decision making. BMC Medical Ethics 7: 12.

[3] Schröder-Bäck, Peter: Principles for Public Health Ethics – A Transcultural Approach. Eubios Journal of Asian and International Bioethics 17(4), 2007: 104-108.

PHGEN II

Systems Biology inspired definitions

A **gene** (originally defined as pangene by Hugo de Vries, with due reference to Mendel) is an irreducible persistently inherited, material factor that co-determines a form or function ('trait') of a living organism. Through scientific discovery it has been established that the material basis of virtually all terrestrial genes is a piece of DNA. Consequently, the word gene and the piece of DNA corresponding to it have effectively become synonymous. Academically speaking however, some genes have the material form of a piece of RNA (in RNA viruses), or even more rarely of a protein (in prions). Functions often correspond with pathways more than macromolecules. Genes correspond with macromolecules, mostly proteins, sometimes structural or functional RNA. DNA that corresponds with a gene needs not be contiguous.

The term **mutation** refers to a change in the primary DNA structure (DNA sequence, i.e. excess, absence or substitution of any of the nucleotides dA, dG, DT, or dC) producing a gene or a chromosome differing from the wild-type. Mutations that are not polymorphisms (see below) are rare in the population. Mutations may be hereditary or acquired. **Hereditary mutations** are inherited from parent to offspring and incorporated in every body cell of the individual, including its germ cells and hence possibly in its progeny. Either parental germ cell or the zygote itself may also have acquired *de novo* mutations. Then, the mutation is present in every cell of the organism including its germ cells. *De novo* mutation can occur also in a germ cell during very early embryonic development resulting in a mosaic cell population present in the gonads only (gonadal mosaicism). Such mutations arising subsequently in a germ cell stand a smaller chance of ending up in the progeny. **Germline mutations** do not affect the person's phenotype, but may be manifest themselves in its offspring.

By definition, **somatic mutations** are mutations in cell lineages that do not lead to gametes. Thereby the mutations do not enter the germline and are not inherited. They arise at any time after postzygotic cell division throughout the whole lifetime of the individual due to the fact that any replicating cell will generate on average a certain number of changes in DNA sequence of a daughter cell in each round of replication. Some somatic mutations are repaired, some persist with the complementary nucleotide inserted into the other DNA strand, and transmitted to daughter cells leading to mosaicism of the individual. These somatic mutations have the effect that somatic cells in an individual are somewhat different from each other genetically. Similarly, somatic cells will differ between monozygotic twins although to a lesser degree than between heterozygotic twins. The degree of difference of the latter will in part depend on the number of replications of the germ cells following the replication of the common precursor stem cell. The term genetically/genomically indentical twins is therefore misleading and should be avoided. Depending on the time of occurrence of a somatic mutation during embryonic or fetal development or later in a permanent cycling somatic stem cell population, the mutated genes may be found in all tissues and organs derived from individual cell lineages or only in a small subset of

differentiated cells. Somatic mutations have strong functional significance in some immune cells, in tumor cells and possibly in some other tissues.

The term **polymorphism** refers to a mutation that occurs at a fairly high frequency in the population and tends to lack adverse effects on the individual under normal conditions.

Genetics is the academic study of how genes determine or influence form or function ('traits') of living organisms. The term is also used to refer to co-determinism of trait and the specific forms (alleles) of a gene. **Epigenetics** refers to a similar determinism by other factors than the base sequence of the DNA in terms of the four canonical nucleotides dA, dG, dC, and dT. To be epigenetic, the determinism should be hereditary for one or a few cell divisions. It may be caused by covalent modifications of chromatin, such as methylation of DNA or acetylation of histones. The generation of modifications that are conserved during cell replication is called **imprinting**. Mechanisms exist that copy such chromatin modifications onto the chromatin of both daughter cells.

The **Genome** is the total set of genes in and of (i.e. excluding the microbiome) an organism that determines the complete set of functions of the latter and hence its maintenance as a species. More precisely it is the set of genes that is transferred through the germ line to the progeny.

The **microbiome** is the total of all the genomes of micro-organisms inside or physically associated with an organism.

The **metagenome** of an organism is the total of its genome and its microbiome.

Genomics is the molecular identification and characterization, in terms of nucleotide sequence, of the genome of an organism.

Functional genomics is the analysis, in terms of molecular identities and amounts, of the expression of the genome of an organism, including all molecular species in between genome and the organism's functions. The expression of a genome occurs though a succession of levels, i.e. the level of mRNA, as studied by 'transcriptomics', the level of proteins, as studied by 'proteomics', the level of metabolites, as studied by 'metabolomics' and the level of functions, as studied by 'physiomics'. The latter is only arguably contained within the class of genomics, as it does not refer to molecular entities but to entities at a higher level of aggregation, such as pathways.

The **transcriptome** is the total collection of RNA molecules that the genome expresses at a certain moment in time and space. In its quantitative form it is represented by the list of numbers in which each number specifies the concentration of an mRNA molecule (or rRNA molecule, or snRNA molecule), and the total list corresponds to the entire active genome. Zeros in the list then represent mRNAs that are present at concentrations below the detection limit. **Transcriptomics** is the molecular analysis of a transcriptome in terms of characterization and quantification.

The **proteome** is the collection of protein molecules that the genome expresses at a certain moment in time and space. In its quantitative form it is the list of numbers that specifies the concentration of each polypeptide chain encoded by a gene. **Proteomics** is the molecular analysis of a proteome in terms of characterization and quantification.

The **metabolome** is the collection of low molecular weight molecules present in an organism, each of which have been processed by one or more of the enzymes present in the proteome. In its quantitative form it signifies the corresponding list of concentrations. **Metabolomics** is the molecular analysis of a metabolome in terms of characterization and quantification. **Metabonomics** is metabolomics used as a tool in drug and biomarker discovery.

The **physiome** is the collection of functions that can be carried out by an organism and are important for its fitness and survival. **Physiomics** is the analysis of a physiome in terms of characterization and quantification of all individual functions.

Omics is any collection of the above, e.g. transcriptomics plus proteomics plus metabolomics plus physiomics.

An **organism** is an autonomous unit of Life, i.e. a network of chemical and physical processes that is able to maintain itself actively (i.e. in free-energy dissipating processes) notwithstanding various assaults it experiences from its environment, and without the dynamic, full time involvement of other organisms. Important maintenance strategies include repair and replication, and generation and survival of progeny ensuring that the network, or a copy thereof, continues to exist. A virus is not an organism, neither is a robot.

> **Distinctions between genomics and systems biology**
> Research has shown that there are no organisms with genomes with fewer than 300 genes and that further removal of these genes removes vitality. This supports the idea that through the expressed genome, functionally important genes are connected to at least 300 if not all genes of the genome. Consequently, no full function or dysfunction of an organism can be determined solely by a single gene, even if a mutation in a gene is the cause of dysfunction in a particular instance is due to a mutation in a single gene. Consequently, there is no such thing as a strict **single-gene** disease and all diseases appear to be systems biology diseases. Each pathway, network, or systems biology disease may be caused by a variety of mutations in genes, or environmental factors, all leading to a similar change in functioning of the pathway. Multiple 'single gene-diseases' may hereby correspond to a single systems biology disease, or to a grouping of such diseases on the basis of their pathophysiology.

Systems Biology is the study of the extent to which and how all functions and functionalities of an organism arise in interactions between its components. It thereby also elucidates how components of a living organism, e.g. genes, contribute to its life by interacting with the other components of the organism. Like genomics, systems biology is comprehensive, i.e. in principle it looks at all functions of the organism and all its components. It does this because virtually all components connect to each other. Functions are here defined in the biological sense, i.e. functionalities that contribute to evolutionary fitness and survival of the species. Because new properties

only arise in interactions when the latter are 'nonlinear', and because they do this only when the nonlinearities are of certain types and strengths, systems biology requires a quantitative experimentation under the actual physiological conditions that is ultimately genome-wide, as well as mathematical analyses in order to deal with the complexity. Consequently, for some the definition of systems biology has become the approach of integrating mathematical and experimental analyses in biology. Systems biology comprises genomics as well as functional genomics; it relates the various ~omics defined above to each other.

Bottom-up systems biology begins with components of organisms, characterizes their abilities to interact with other such components, and then examines how the functional properties that are important for the organism as a whole, emerge in the interactions. Ultimately it aims for understanding at the level of the entire organism, but this in terms of component properties and properties emerging from component interactions.

Top-down systems biology begins with an omics study, i.e. it looks at the complete set of molecules at a certain level of gene expression (e.g. at the transcriptome) and then tries to recognize patterns of change when conditions vary in terms of time, space or external environment. It then tries to induce from the observed patterns empirical principles of how collective functions emerge from interactions. Ultimately it may aim for understanding down to the molecular level.

Middle-out systems biology begins at any level between the functioning whole organism and a few components. It addresses emergence of function at that level from interactive properties just below or at that level.

Integrative systems biology integrates top-down, middle-out and bottom-up systems biology methodologies with each other and with all other relevant sciences.

Systems Medicine is the systems biology of the human as it pertains to disease.

A (single) **gene disease** is a disease that is fully attributable to (caused by) a mutation in a single gene. In a looser meaning it is a disease that is significantly correlated with such a mutation.

A **genetic disease** is a disease that is fully attributable to (caused by) mutations in one or a number of genes. A difference with network diseases is that no reference is made to relations of the genes through their encoding of proteins being positioned in common intracellular networks. There is debate whether genetic diseases exist at all, for instance because most diseases are *also* caused by environmental factors, hence not exclusively by mutations. However, there may well be diseases that always appear when the individual has a certain number of mutated pieces of DNA inherited from its parents. Genetic diseases are the (probably large) subset of inherited disorders that are not due to epigenetics, i.e. for which the mutations reside in the primary DNA sequence.

A **familial disease** is a disease with a higher than usual frequency of occurrence in a family of individuals, but which has not been demonstrated to be an inherited disease.

A familial disease could be caused by environmental factors to which the specific family is subject.

A **genomics disease** is a disease that is caused by a loss of the connectivity of the genome as such, such as due to chromosomal rearrangements in certain tumors. In a much wider sense the words 'genomics disease' refer to a rupture of the genome as unit as such. When this is taken to comprise all gene-gene interactions through the expressed genome, this becomes similar to a systems biology disease.

A **pathway disease** is a disease caused by a failure in a pathway, i.e. a subnetwork of the cell that serves a rather defined function.

A **Systems Biology disease** is a disease caused by a failure in functioning of networks in the living organism, i.e. it is due to a breakdown in interactions between components that lead to the normal functioning of an organism.

A **network disease** is something in between a pathway disease and a systems biology disease.

Medicine is the set of science-based activities directed at understanding, diagnosing, managing and curing diseases.

In **generic medicine,** disease is classified on the basis of the type of malfunctions that have been observed in large numbers of human individuals and the results of corresponding medical-biochemical, physiological and imaging investigations. A patient is diagnosed in terms of suffering from any of the standard disease types. Patients diagnosed as suffering from the same disease are treated in essentially the same way. Their therapy is virtually irrespective of individual characteristics of the patients that are not directly a consequence of the disease.

Cohort-based medicine divides patients of a certain disease into two, or a small number of, cohorts on the basis of criteria that are not directly related to the malfunctions of the disease. Rather, these criteria relate to molecular or other characteristics that appear to correlate with the success rate of the therapy of the disease. The therapy then depends on the cohort a person has been placed into.

Personalized medicine decides on therapy both on the basis of the disease diagnosis and on the basis of characteristics of the individual patient that may or may not be caused by the disease. An obese patient may be given a larger dose of a hydrophobic anti-cancer drug than a lean person. The dosage of a cytostatic drug may be made dependant on the psychological state of the patient. Although personalized medicine adjusts to the characteristics of the individual, it still does this in fairly generic terms, and will treat similar patients identically.

(Truly) individualized medicine is not yet practiced. It is based on the realization that with the exception of identical twins, all human individuals are different in their genome (approximately 0.1% difference in coding sequence between the germline cells of two individuals) and consequently in the molecular networks that co-determine their cellular and hence somatic overall behavior (because of the influence

of environment, this is also true between identical twins). Since behavior in one area of the network will depend on the rest of the network, no two diseases in different individuals will be the same and consequently also the therapies between each two different individuals should be different according to the paradigm of truly individualized medicine. Although individualized medicine may appear to be a more extreme form of cohort-based medicine, the fact that its cohorts essentially have a size of 1, completely changes the statistics around clinical trials and the strategy of drug development.