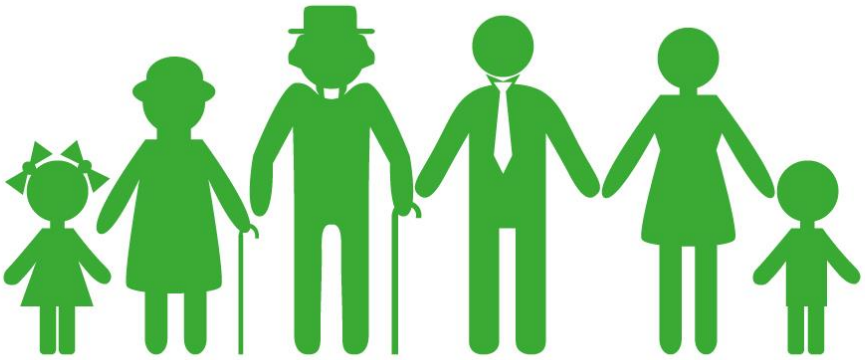

DIET, CANCER AND HEALTH

NEXT GENERATIONS

Diet, Genes and Environment

Anne Tjønneland
Anja Olsen
Jytte Halkjær
Ole Raaschou-Nielsen
Per Guldberg

July 2014

Indhold

Summary

The “Diet, Cancer and Health” (DCH) cohort includes 57,053 Danish citizens. At enrollment in 1993-1997, all participants were aged 50-64 years and had no previous cancer diagnosis. Detailed information on diet, lifestyle and anthropometric measures was recorded, biological material (blood, urine, fat tissue and toe nail clippings) were collected, and all participants have continuously been followed by linkage to the national population and health registries. This cohort has formed the basis for numerous national and international prospective studies on the impact of diet, metabolic characteristics, lifestyle, environmental exposures and genetic factors on risk of cancer and other health conditions.

The overall objective of this proposal is to extend the existing DCH cohort by including biological children and grandchildren of members in the original cohort. Registry data will be acquired on all

children, their spouses and their children (279,176 individuals), and a large subsample (n=187,564 individuals) will be invited to participate in a baseline examination and collection of biological material. With an expected participation of 30%, this will result in a second and third generation cohort of around 50,000 Danes. The biological samples will be stored in the Danish National Biobank and made accessible to a broad research community.

The resulting three-generation cohort will allow in-depth trans-generational studies of the role of genetic, epigenetic, microbiomic, environmental and socio-economic factors and their complex interactions in the pathogenesis of multiple cancer and other noncommunicable diseases. Similarly, the cohort will constitute a unique source for studies of biomarkers predicting such common disorders.

Rationale and Objectives

Enabling the individual to make informed choices on health behavior requires a profound understanding of the genetic predisposition to specific health conditions as well as knowledge of lifestyle and additional environmental exposures that may affect disease risk. Such information is usually obtained from epidemiological studies. Cross-sectional studies have, however, major limitations when it comes to distinguishing between lifestyle and genetic predisposition. In contrast, prospective cohort studies provide the opportunities to achieve both lifestyle information and biological materials, which are unaffected by the disorders that the study participants eventually may contract. In this research proposal the Diet, Cancer and Health – Next Generations (DCH-NG) will take advantage of the strengths of the prospective cohort study design and further add the trans-generational dimension.

The original DCH cohort includes 57,053 Danish men and women, aged 50-64 years at the time of enrollment (1993-1997). All participants underwent extensive assessments at baseline, including detailed information on lifestyle (e.g. diet, physical activity and smoking), and provided samples of blood, urine, fat tissue and nail clippings. Since its establishment, the cohort has been used for international research of high quality, and numerous studies have been conducted on lifestyle, and on genetic and environmental risk factors for disease. Thus, until now, the cohort has contributed data in more than 600 scientific publications in peer-reviewed journals in addition to multiple doctoral dissertations, PhD degrees and master theses.

In DCH-NG biological children and grandchildren of the DCH participants will be invited to participate in a study similar to the one in which their mother, father or grandparent participated 20 years ago. This will create a 3-generation cohort. DCH-NG is estimated to include 50,000 new participants from whom diet and lifestyle data as well as biological material will be collected, including DNA, serum, plasma, erythrocytes, urine, saliva and feces samples.

Therefore, the overall aim of the present initiative is to establish the DCH-NG as a unique resource – open for the broad research community - for future trans-generational studies of the pathogenesis of multiple cancer and other noncommunicable diseases. Likewise, the cohort is projected to be most valuable in the search for biomarkers predicting such common disorders.

The exploitation of the information collected from DCH-NG will be accelerated by the recent advances in technologies able to generate, analyze and integrate high-content data, including data from next-generation sequencing technologies, and different "omics" platforms for high-throughput analysis of the genome, epigenome, transcriptome, proteome, metabolome and microbiome.

The DCH-NG cohort

Children (hereafter named G2) and grandchildren (hereafter named G3) of DCH participants (hereafter mentioned G1) compose the relevant study population, together with the parent of G3 NOT being linked to DCH through G1 (hereafter called G2-A).

We have identified 279,176 persons in G2 and G3. These individuals together with the 57,053 participants in the original DCH cohort (G1), constitute a cohort only intended for registry-based research and will be the basis for selection of participants to the DCH-NG cohort with extensive collection of data on diet, lifestyle and biological samples. For details see **Figure I** and **Table I**.

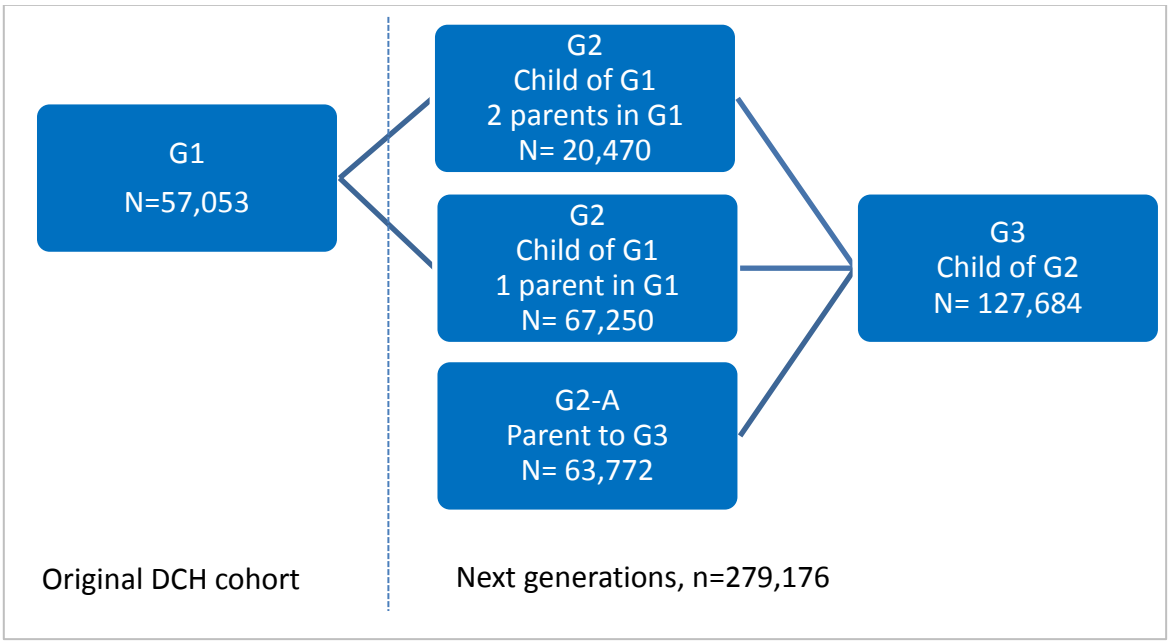

Figure I. G2, G3 and G2-A persons identified in the CPR registry.

The selection of participants for inclusion into DCH-NG will secure best possible inclusion of participants with familiar link. G2 is considered as the index person meaning that if G2 meets the inclusion criteria, he/she as well as his/her family (G3 and G2-A) will be invited to participate in the study.

Among the 279,176 G2 and G3 persons we will limit invitations to participate in the DCH-NG cohort to:

For G2

1. Persons alive at time of invitation
2. Persons identifiable and contactable in the Central Population Registry (CPR)
3. Persons being at least 18 years of age at inclusion

For G3

1. Persons whose G2-parent meets the inclusion criteria
2. Persons alive at time of invitation
3. Persons identifiable and contactable in CPR
4. Persons being at least 18 years of age at inclusion

For G2-A

1. Persons being a parent to a G3 together with a G2 that meets the inclusion criteria
2. Persons alive at time of invitation
3. Persons identifiable and contactable in CPR
4. Persons being at least 18 years of age at inclusion

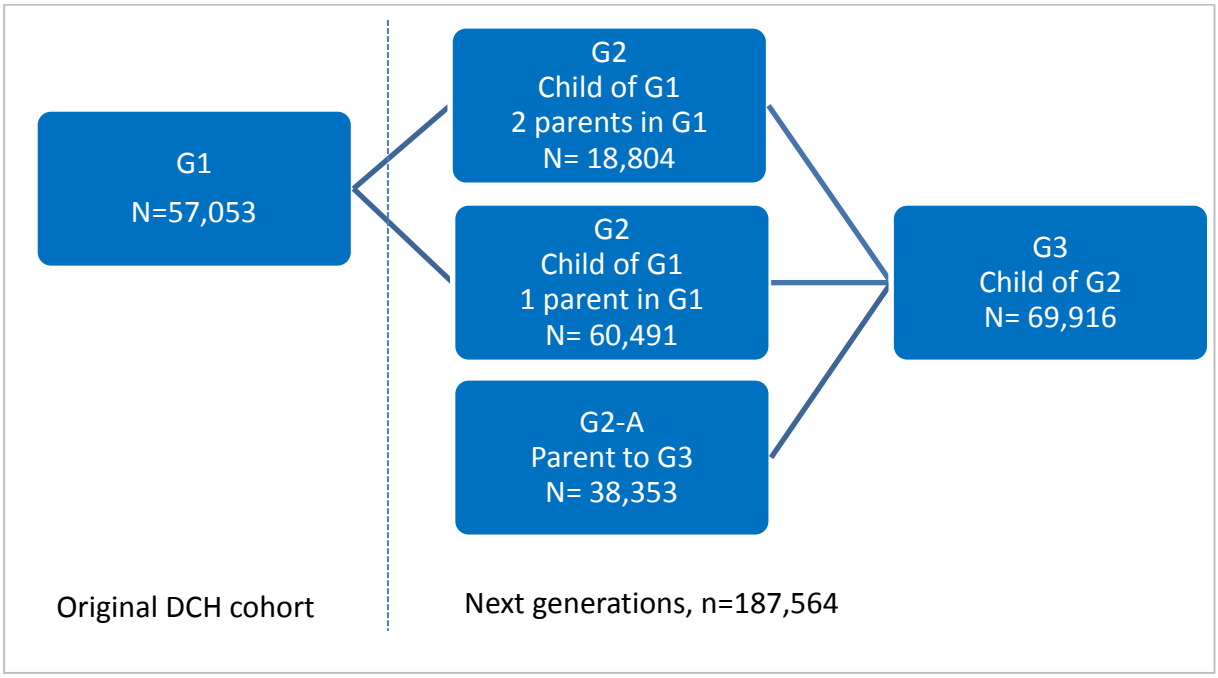

Figure II. G2, G3 and G2-A identified in the CPR registry and meeting the inclusion criteria.

A total of 187,564 persons fulfill all criteria for invitation to DCH-NG. For details see **Figure II** and **Table I**. An unknown number of these will thus be excluded due to non-participation of their relevant G2 or G3. Provided a participation of 30 %, about 56,000 participants will be enrolled into the new DCH-NG

cohort. When G1 was invited to the original cohort, the participation was 35%. We expect a participation rate in the same magnitude or a little lower for the next generations.

Table I. Overview of the three generations according to sample size, age at enrolment and birth year.

Total number (for registry studies)	**Count**	**Age range**	**5-95%**	**Birth year**	**5-95%**
DCH (G1)	57,053	50-65	50-65	1929-1947	1931-1945
Children of DCH (G2)	87,720	1-65	33-55	1947-2011	1957-1979
Grandchildren of DCH (G3)	127,684	0-48	2-29	1964-2012	1983-2010
Other parents to grandchildren of DCH (G2)	63,772	17-88	34-59	1924-1995	1953-1978
Contactable*	**Count**	**Age range**	**5-95%**	**Birth year**	**5-95%**
DCH (G1) (will not be contacted again)	46,688	67-85	69-82	1929-1947	1932-1945
Children of DCH (G2)	79,295	15-67	35-57	1947-1999	1957-1979
Grandchildren of DCH (G3)	69,916	15-49	15-32	1965-1999	1982-1999
Other parents to grandchildren of DCH (G2-A)	38,353	31-85	42-62	1929-1983	1952-1972

*see the inclusion criteria in the text

Strengths of DCH-NG

As outlined in the general objective the DCH-NG resource (biological samples and measurements and registry-based phenotype information) will be the basis for future scientific exploitations of the complex interplays between individual genome and microbiome variations on the one side and non-genetic individual behavior and exposures on the other side in the pathogenesis of cancer and other multi-factorial disorders. For these purposes the DCH-NG will have several strengths:

Prospective samples. The pre-diagnostic biological samples are a valuable resource addressing the concern of post-diagnostic biomarkers in e.g. case-control studies. In such studies the concern is that the disease might have caused changes in the biomarker among the cases rather than interpreting different biomarker values between cases and controls as associated with the causal pathway to disease, i.e. the problem of inverse causation.

Age of the second generation. It is a major advantage of the DCH-NG study that the second generation already has reached an age where cancers and several other diseases usually occur. In newly established birth cohort studies, which also include more generations, it takes several decades before the more frequent diseases such as cancer, cardiovascular diseases and diabetes 2 develops in the second generation. **Figure III** illustrates the distribution of possible participants according to birth year and age at enrolment.

Record linkage with Danish registers. Danish nationwide population-based registers facilitate follow-up for cancer and a wide range of other diseases among the participating cohort members. It also

facilitates "extended" register-based studies of all individuals in three generations (see **Figure I and Table I**), including those invited but choosing not to participate or not fulfilling the inclusion criteria in the new active data collection. Record linkage will be made to the Danish National Patient Register, the Danish Cancer Registry, the Danish National Prescription Register and other registries in order to follow up for disease outcome for all offspring from the DCH study. It will also facilitate studies on the use of health care services by subgroups and analysis of non-response bias (1;2). Further, clinical databases and biobanks of cancer and other diseases in Denmark can provide detailed diagnoses and biological measurements for characterization of cases developed in the DCH-NG cohort in addition to what is available in the nationwide population-based health registers.

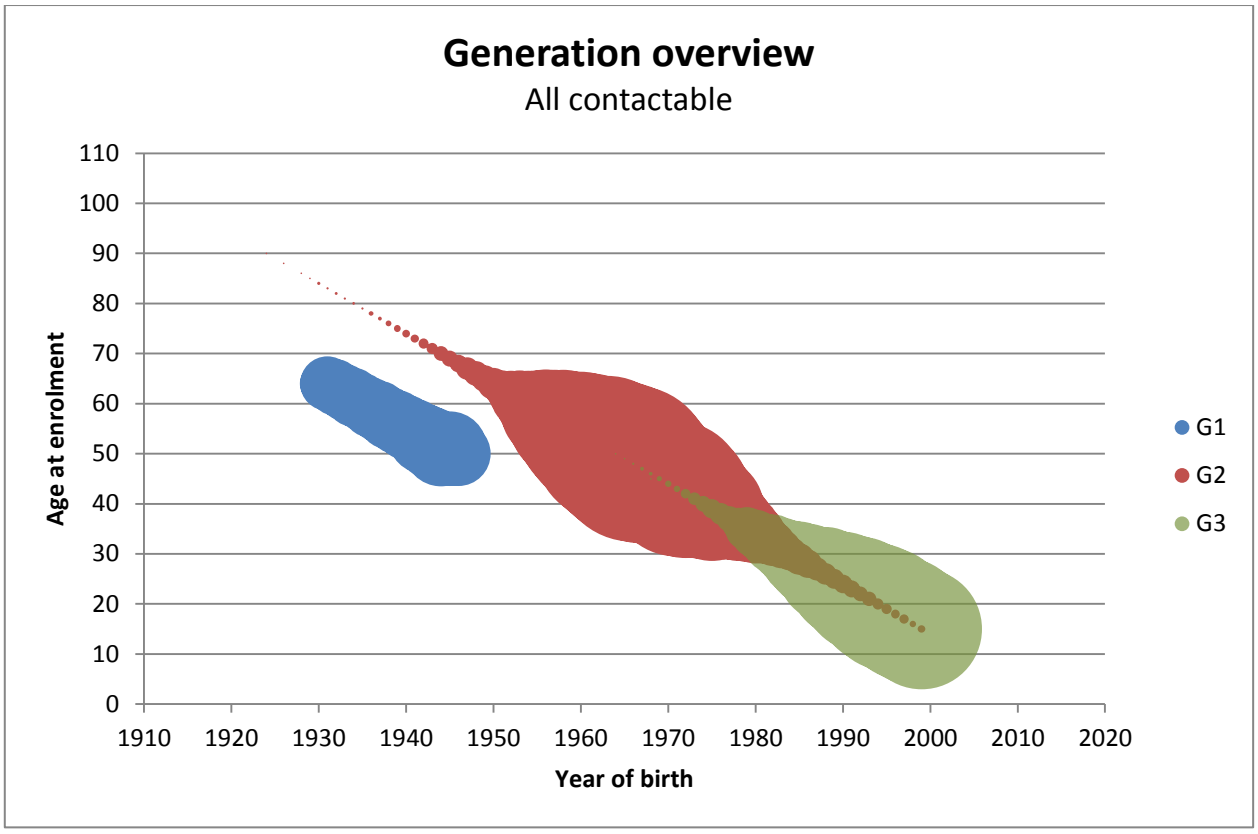

Figure III. Distribution of possible participants according to birth year and age at enrolment. "Contactable" refers to possible participants (G2, G2-A and G3) found in the registry and fulfilling the inclusion criteria for participation in DCH-NG. Enrolment for G1 was 1993-1997; enrolment for G2 and G3 is defined as December 2012.

Trio design. Previous genome-wide association studies (GWAS) have identified >200 genetic loci that are associated with risk of a range of common human diseases, including many cancer forms (3). However, each of these loci has shown weak associations to disease and thus most of the supposed heritability of common diseases remains unexplained. Possible explanations for this 'missing heritability' include limited statistical power of individual studies, lack of refined phenotype ascertainment, population heterogeneity, and parent-of-origin-specific effects. It has become increasingly clear that family-based approaches may circumvent many of the limitations of population-

based studies, thus increasing the probability of verifying causal associations. One important example of a family-based design is the haplotype relative risk (HRR) method (4), where nuclear families with an affected individual and his or her parents ("trio design") are sampled. The design of the DCH-NG cohort will generate a large number of trios (estimated 20,000 by 30 % participation) and thus may be a useful resource for accelerating the discovery of genetic predisposition in polygenic traits.

Research Fields and Perspectives

The DCH-NG cohort and biobank will constitute a unique resource for studies into transmission of biological, dietary, behavioral, socio-economic and environmental factors through generations. Most of such transmissions probably have both genetic and non-genetic components and studies based on the DCH-NG cohort will contribute important novel and original knowledge in this field. Some potential research areas are:

Heredity. The DCH-NG study will facilitate study of the transmission to next generations of lifestyle, environmental exposures, and diseases. Exploitations within this research field will address questions like:

- To which extent are lifestyle, environmental exposure and disease patterns transmitted to next generations?
- Which proportion of such transmissions can be explained by individual genome variation and which proportion by non-genetic factors such as health behavior and socio-economic status?

Rare, intermediate-penetrant genetic variants and copy number variation. Previous family studies have identified rare disease-associated genetic variants with high penetrance. In addition, genome-wide association studies (GWAS) have identified a high number of common low-penetrant variants. Due to the limited statistical power of previous array-based genotyping studies, however, relatively rare genetic variants with intermediate penetrance and various forms of copy number variations have remained undiscovered. The DCH-NG study may, through its trio design and by applying targeted and/or whole genome/exome 3-generation sequencing with high coverage (>50X) provide sufficient statistical power to identify novel genetic risk factors within this "black box".

Gene-environment interactions. A major focus will be studies of interaction between genome variation and lifestyle taking advantage of the 3-generation approach, which allows inclusion of previous generations' risk factor and disease profiles as determinants of corresponding risk profiles in the next generations. Similarly, DCH-NG provides novel opportunities for bioinformatics-driven discoveries of molecular pathways and gene networks that mediate various environmental effects.

Trans-generational epigenetics. Recent work has suggested that inheritance of phenotypic traits can be epigenetic, i.e., transmitted from one generation to the next by modification of the genome without altering the primary DNA structure. However, the relative importance of genetic and epigenetic inheritance is unknown. The trio design of DCH-NG will allow a detailed investigation into the

inheritance of epigenetics such as DNA methylation and histone modifications in lymphocytes and the associations with specific health conditions.

Saliva and gut microbiomics. Several hundred trillions of microbes (primarily bacteria) live in, on and around us all the time, contributing approximately 2 kilograms to our body mass. Most of them are essential for our health and live in equilibrium with our bodies. The vast majority of these bacteria live in our guts and breaking this balance may have various consequences such as some forms of cancer and a series of inflammatory, immune and metabolic disorders. Knowing which microbes live in our saliva and gut is hypothesized to lead to better, personalized diets, early diagnosis, and treatment of multiple common diseases. DCH-NG will be the basis for a series of quantitative saliva and gut microbiome studies in relation to elucidate the pathogenesis of multiple cancers and inflammatory, immune and metabolic disorders.

Biomarkers for early detection of disease. The large collection of pre-diagnostic blood, urine, feces and saliva samples in the DCH and DCH-NG cohorts will provide a valuable resource for studies investigating the potential of early and non-invasive detection of cancer and other diseases, using a variety of powerful 'omics' technologies that can detect tumor-associated biomarkers (DNA mutations, DNA methylation, microRNA, microbiomics, metabolomics etc.). These novels 'omics'-identified biomarkers will be critical elements to be integrated in the development of algorithms for disease risk assessments.

De novo mutations. De novo mutations refer to new mutations originating in germ cells or in the fetus. Complete trios are required for exploitation of the rate and nature of de novo mutations and DCH-NG will consequently be a unique resource for research within this area. In circumstances where it is not possible to construct complete trios, we will take advantage of in silico genotyping to infer the sequence of the ungenotyped father/mother.

Metabolic parameters in generations. Metabolic parameters of relevance for disease risk, e.g. obesity, insulin resistance, high cholesterol levels and inflammatory markers are known to cluster within families. A number of genetic variations related to these factors have been identified, but only explain a minor part of the observed heredity. A generation study will allow a combination of lifestyle and genetic factors related to metabolic parameters and thereby provide a deeper insight into the combined effects of these factors.

Data collection

Invitation procedures

Two investigation centers will be established in Copenhagen and Aarhus to handle contact, enrollment and the physical examinations. Eligible participants have been identified in the CPR and will be contacted by letter and asked to participate. Non-responders of this letter will be contacted by an additional letter. For questions from participants or potential participants a hotline will be established. The children of the original cohort members (G2) are the index persons meaning that they will form

the basis for invitation of participants and securing a familiar link as described earlier. If interested, the participant will sign up at the study homepage (www.kkhng.dk) and get access to a personal web profile that is generated upon enrolment. Registration and informed consent will be signed by help of the Danish digital signature 'NemID'. From the personal profile the participant can access the online booking system for the physical examination and fill out the internet-based questionnaires. The internet platform will have an online support function.

Questionnaire information

Dietary data will be collected via a web-based, comprehensive food frequency questionnaire, and repeated 24-hour recalls. Further, a questionnaire on lifestyle will cover exposures such as smoking, physical activity, reproductive variables (pregnancies, children, lactation duration, use of contraception), education, sun exposure, and several other health aspects. The questionnaires will preferably be filled out at home by participants with computer access and for participants without a computer, the questionnaire will be completed by telephone interview. The computer program will not accept missing answers. The questionnaires will resemble the original Diet, Cancer and Health questionnaires, but particularly with regard to diet, modifications are needed to account for new foods and dietary patterns in a younger generation. The Danish National Birth Cohort, which also used the Diet, Cancer and Health food frequency questionnaire and software Foodcalc, has already made such modifications and validations. This questionnaire is available in an internet-based version, which, with minor adaptions, will be used.

In addition to the information obtained by questionnaire the usability of novel techniques such as smartphone applications (apps) for assessment of physical activity will be explored. Concurrently with the fast development in this field, the app may further develop during follow-up to include other interesting areas.

Physical examination

Body weight, body height and waist/hip circumference will be measured by trained personal after the same standardized procedures as already applied for G1. In addition to usual anthropometry, participants will also be measured by a body analyzer using bioimpedance to obtain a better and more accurate measure of body composition such as the amount of total and bodysite specific adipose tissue and muscle mass. Blood pressure, pulse and heart rate variability will also be obtained after standardized procedures.

To get a more objective measure of fitness the participants will be asked to take a step test developed and validated in Denmark. The test is a submaximal test and rated as a safe and feasible alternative to the more time-consuming watt-max test as a method for estimation of VO_{2max} in large adult population-based studies (5)

Biological samples

Experienced/trained staff will take and handle biological samples of blood separated into serum, plasma, buffy, and erythrocytes and a 2.5 ml PAX tube for RNA as well as urine and saliva under strictly

standardized conditions. Furthermore, equipment and instruction for standardized collection of feces samples at home will be handed to the participant. The frozen samples will subsequently be returned to the clinic. **Table II** summarizes the types, amounts and possible application of the collected materials. Collection and storage of the samples will be done according to state-of-the-art procedures to ensure optimal usability of the material for the future.

Table II. Biological material

Biological material	Amount	Examples of use
Serum	4 ml	Vit D, enterolactone, hormones
Plasma (EDTA and Lithium-Heparin)	4 ml each type	Vitamines, C-peptides, PSA, metabolomics
Erythrocytes	1 ml	Hem-adducts (e.g. acrylamide)
Buffy coat	1 ml	DNA extraction for DNA sequencing and SNP analyses
Whole blood (RNA tubes)	2.5 ml	RNA, microRNA
Spot urine	4 ml	Metabolomics, trace elements, heavy metals, DNA
Saliva	2 tubes (1 plain, 1 with RNAlater)	Microbiota, RNA, DNA, microRNA
Feces	1 tube	Characterization of the intestinal microflora; shotgun metagenomics, DNA
Plasma (Li-Hep)	2-3 ml	For upfront analyses (see below)
Whole Blood (EDTA)	2-3 ml	For upfront analyses (See below)

Upfront analyses

In addition to the blood samples collected for storage in the biobank for future research, whole blood and lithium-heparin (Li-Hep) blood samples will be taken for upfront analyses to be performed immediately after the visit in the study center. From whole blood *hemoglobin A1c* as a measure of average long term glucose levels will be measured, and from Li-Hep plasma total, HDL and LDL cholesterol, triglycerides, high sensitive CRP and creatinine will be measured. The values will be provided to the participants if interested, and data will be included in the research database for immediate use in research projects.

Reproducibility and validity

For estimation of biomarker stability, 1000 randomly selected participants will be asked to complete supplementary questionnaires and provide further biological samples 6 and 12 months after their baseline examination. The number of repeated samples needed, will vary according to the specific marker, but 1000 participants with two extra samples are regarded as sufficient for most foreseeable situations.

Validation studies will be conducted regarding the information collected by questionnaires.

Ethics

Before initiation of the project, a detailed description of DCH-NG will be presented for the relevant ethical committee and the Data Protection Agency and no potential participants will be contacted before this is approved. Participation is conditioned by signed informed consent. For future use of DCH-NG permissions will be obtained from ethical committees and the Data Protection Agency according to Danish legislation.

Considerations regarding statistical power

DCH has proven sufficient power to be basis for a large range of epidemiological and genetic-epidemiological studies. The DCH-NG will have a larger sample size and though several of the next generation participants will be young, and thereby have a low incidence of many diseases, a high fraction has already reached an age where many of the relevant disease starts to occur with pertinent rates.

From a statistical point of view the DCH-NG cohort will provide an important resource with the opportunity of designing powerful studies to evaluate shared genetic and environmental influences on different outcomes. Designs that involve families as well as population-based sampling allow investigation of both genes and environment, separately or together and allow valid inference to the population. The family based designs have advantages over traditional population-based designs, as they are robust against population admixture and stratification and allow both linkage and association to be tested for. Furthermore the fact that family-based designs contain both within- and between-family information has substantial benefits in terms of multiple-hypothesis testing, especially in the context of whole-genome association studies. All in all the size and structure of the DCH-NG cohort will permit feasible studies investigating the effects of both genes and environment for many outcomes.

For low frequency variants (minor allele frequency 1-5%) and rare variants (< 1%), the statistical power in the DCH-NG cohort alone is limited when using single variant-based statistical tests. However, the application of various validated collapsing methods may improve power by combining all variants within a gene to be studied as a single analytical unit. In addition, prioritization of variants by the genome-wide analysis of phenotypic variance and pathway analyses, and prioritization of environmental factors by the calculation of variance explained by gene-lifestyle interactions will increase statistical power by reducing the burden of multiple testing. Additional variants may be identified by international collaboration with comparable cohorts and performing meta-analyses.

Scientific quality

This protocol has been drafted in collaboration with strong national scientific environments, with specific contributions from the persons mentioned in the National Scientific Collaborative Group (**Figure IV**). Further effort to ensure the best possible scientific quality of the proposed cohort study has been taken. In the initial phase in 2013, national as well as international experts and stakeholders

were invited to a symposium with the purpose of optimizing the type and methods of data to be collected.

Organization, management and timeline

The DCH-NG will be coordinated from the research UNIT *Diet, Genes and Environment* at the Danish Cancer Society Research Center headed by Project Leader, MD, PhD, DMSc Anne Tjønneland. Group leaders within the UNIT, Professor, PhD, Per Guldberg and Senior Researcher, PhD, Ole Raaschou-Nielsen will be involved in planning the projects related to genetics and environmental risk factors, respectively. Group leader within the UNIT, Senior Researcher, PhD, Anja Olsen will be involved as senior scientific advisor (see **Figure IV**). A project group with representatives of the different research areas and the day-to-day running will be established and headed by Project Manager Jytte Halkjær. The Project Manager, in collaboration with the Project Leader, will be responsible for the scientific and technical project managing and monitoring, including communication within the project and project visibility and distribution of knowledge to recipients outside the network. They will refer to the Board of Directors of the Danish Cancer Society regarding approval of budgeting and accounts. The DCH-NG database, when established, will serve as an important scientific infrastructure to be used by the broad research environment in Denmark and worldwide. The overall data collection is therefore planned in close collaboration with several key scientific collaborators and research institutions in Denmark.

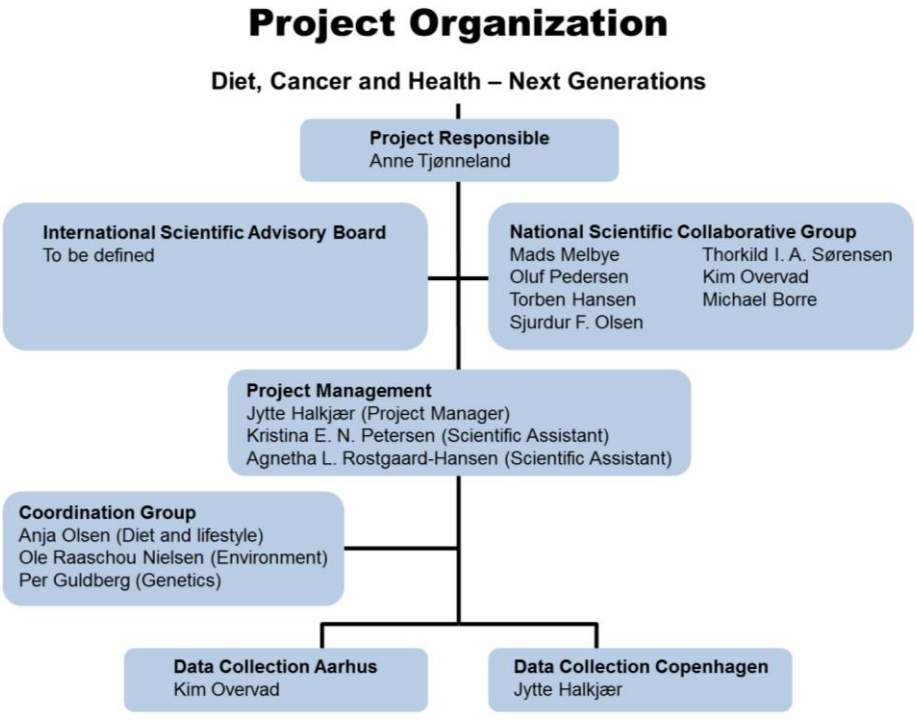

Figure IV. Organization and management.

The National Scientific Collaborative Group. A National Scientific Collaborative Group will be responsible for approving of scientific work plans and strategic issues beyond these work plans.

The Danish National Biobank, Denmark will be represented by Professor and Director Mads Melbye, MD, DMSc from Statens Serum Institute (SSI). The collected biological material from DCH-NG will be stored in the Danish National Biobank (www.biobankdenmark.dk), established at SSI.

The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen (www.metabol.ku.dk) will be represented by Director Oluf Pedersen, Professor Torben Hansen and Professor Thorkild I.A. Sørensen.

Oluf Pedersen MD, DMSc, Professor of Molecular Genetics and Metabolism at Faculty of Health and Medical Sciences, University of Copenhagen and Director of both Novo Nordisk Foundation Center for Basic Metabolic Research (www.metabol.ku.dk) and Lundbeck Foundation Center for Medical Genomics (www.lucamp.org). Oluf Pedersen has a profound expertise and experience in basic and large-scale genomic discovery through next-generation sequencing and array based genotyping, and genetic-physiology and genetic-epidemiology.

Thorkild .IA. Sørensen MD, DMSC, is Director at the Institute of Preventive Medicine, and Professor of Metabolic Epidemiology and a well-known expert within obesity research, genetics and metabolic diseases as well as metabolic genetics.

The National Birth Cohort „Better Health for Mother and Child" (www.dnbc.dk) will be represented by Adjunct Professor at Harvard School of Public Health, MD, MSc, DMSc, Sjurdur Frodi Olsen, who is an expert within the field of early exposures and disease. Sjurdur Frodi Olsen will be a central partner due to the utilization of questionnaires and exchange of data with the National Birth cohort.

Professor of Epidemiology, Kim Overvad MD, PhD, Aarhus University will also be part of the scientific group, ensuring that the data collection will be in line with the newest scientific development within the field. Furthermore, he will be involved in the day to day management of the study clinic in Aarhus.

Close collaboration will also be established with Professor, MD, DMSc Michael Borre, as a representative of the Danish Multidisciplinary Cancer Groups.

The International Scientific Advisory Board. An International Scientific Advisory Board will be established to secure state-of-the-art data collection and storage for generation studies. The composition will, before establishment, be approved by the Board of Directors and the Scientific Council (KBVU) of the Danish Cancer Society.

Decisions on use of the data and biological material for defined scientific studies, once the data collection has been finished, will be the responsibility of an established Scientific Board for DCH-NG. The composition of this board will be decided on in close collaboration with the Project Leader and the

already existing board for the *Diet, Cancer and Health* study, consisting of national experts within the fields of epidemiology, nutrition, genetics, biostatistics and molecular biology.

To get access to data, an application with a synopsis of the proposed study will have to be presented and approved by the Scientific Board.

Progress meetings for the National Scientific Collaborative Group are planned annually and telephone meetings will be held every second month to secure overall project progress, while the International Scientific Advisory Board will have meetings yearly. The proposed high level timeline of data collection is outlined in **Figure V** (following page).

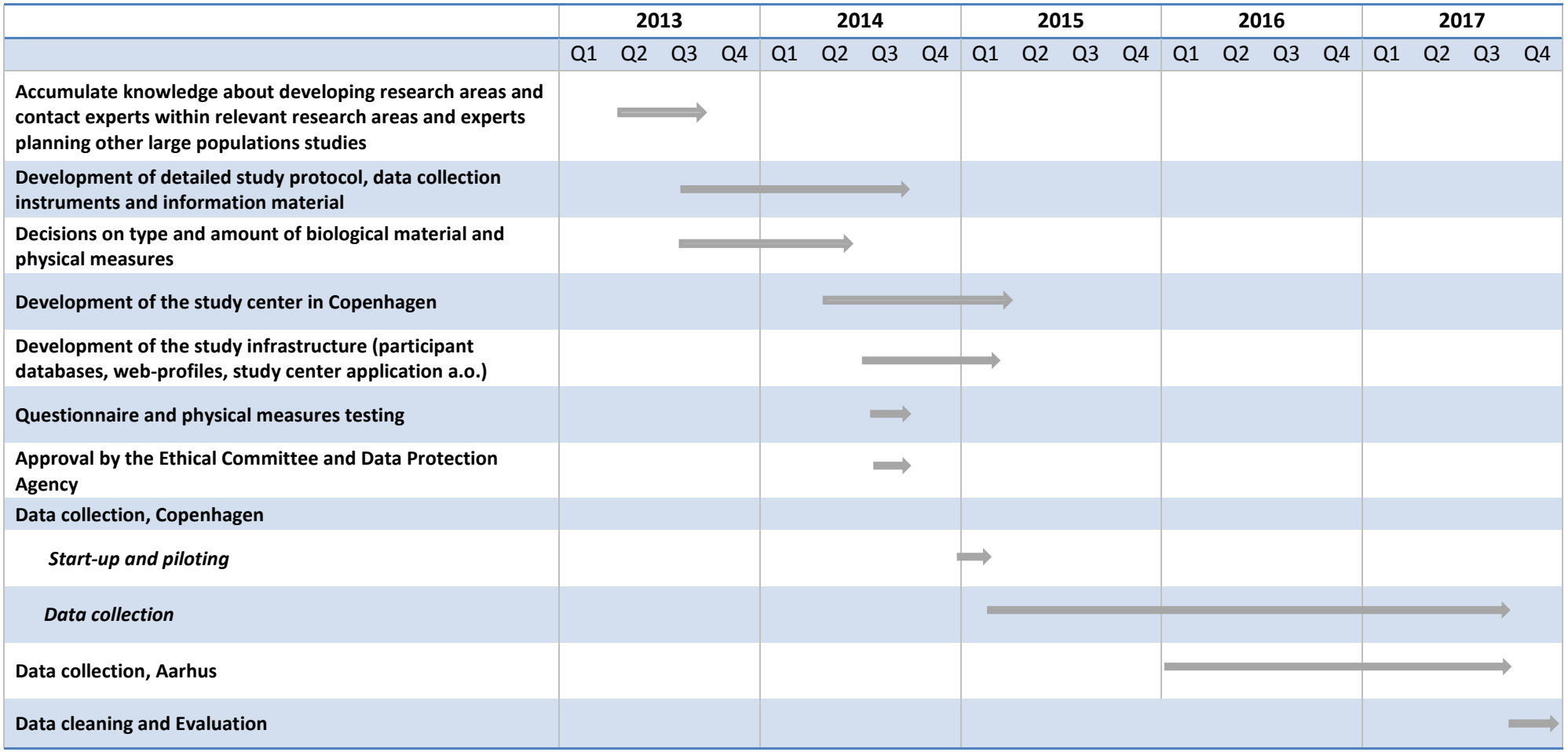

Figure V. Time line (High level)

Reference List

1. Davidsen M, Kjoller M, Helweg-Larsen K. The Danish National Cohort Study (DANCOS). Scand J Public Health 2011;39:131-5.

2. Christensen K, Kyvik KO, Holm NV, Skytthe A. Register-based research on twins. Scand J Public Health 2011;39:185-90.

3. Hindorff LA, Sethupathy P, Junkins HA et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106:9362-7.

4. Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 1987;51:227-33.

5. Aadahl M, Zacho M, Linneberg A, Thuesen BH, Jorgensen T. Comparison of the Danish step test and the watt-max test for estimation of maximal oxygen uptake: the Health2008 study. Eur J Prev Cardiol 2013;20:1088-94.