2 SCIENTIFIC BACKGROUND

The following chapter is considered to illustrate substantial background knowledge of telomeres. It is aimed to provide information on structural components and the decisive role of telomeres in aging and associated diseases as well as on methodological methods applied in this thesis for a better understanding of presented projects.

2.1 BIOLOGY OF AGING

Age in general is defined as the length of time that an individual has lived. Chronological age is a measure of individual's age based on the calendar date of birth which is usually going hand in hand with a certain stage of mental and physical development with the same progression rate in each individual [1]. In contrast, biological age includes the physiology status. It is a result of alterations, progressive degeneration, loss of function and loss of regenerative capacity of cells, tissues and the whole organism. This deterioration is attended by increased frequency of disease occurrence and does not proceed in the same rate in each individual [2]. Although several risk factors of aging and therefore a higher rate of biological aging are already known [2, 3], there is still a challenge to elucidate the principal tasks of phenotypical healthy aging and longevity by prolonging the healthy lifespan and preventing from diseases simultaneously. Respective to its complexity numerous theories on aging are available [4]. But it is proposed that extension of lifespan can be achieved through a favorable genetic constitution that increases the DNA repair mechanisms, reduces oxidative damage and apoptosis [5, 6]. Therefore by preserving genetic integrity, telomeres might have an important impact in regulation of aging process [7, 8].

2.2 BIOLOGY OF TELOMERES

The word telomere originates from the Greek words "telos", meaning end and "meros" meaning component. First telomeric DNA was identifies in 1978 [9]. Telomeres describe regions of random short repetitive nucleotide sequences at the end of eukaryotic linear chromosomes. The sequence motif as well as the quantity of these repeats can be different between species. But the principle task is consistently the sustainment of chromosomal integrity [10].

The number of nucleotides varies not only between species but also between tissues, cell types and chromosomes within one individual. Human telomere length is estimated to comprise about 5-15 kb. The human sequence motif is 5'-TTAGGG-3'. Telomeric DNA is a double-stranded structure which ends at the very end in a 3'-single-stranded G-rich overhang (150-200 bp) [11]. This overhang forms the so-called telomere loop (T-loop) a double-stranded structure occurring by looping back and annealing of the single strand to the double-stranded hexamer repeat. At the distal end of the T-loop, the rest of the single-stranded overhang is attached to a region of the double strand DNA. As a consequence, a triple-stranded structure called displacement loop (D-loop) is formed (**Figure 1**)[12, 13].
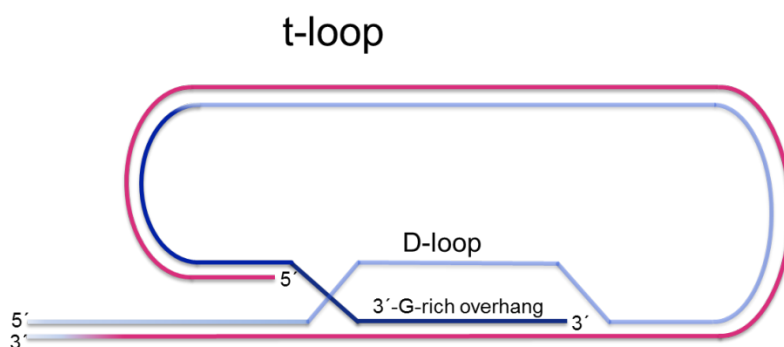
Figure 1: Structure of telomere; T-loop: telomere-loop, D-loop: displacement-loop, 3´-G-rich overhang: 3´-guanine-rich overhang. Figure adapted according to De Lange [12].

This terminal DNA structure is stabilized by a protein complex, which is characterized by several evolutionary conserved regions. By interacting with the telomeric DNA various telomere specific proteins are responsible for an accurate and proper function of telomeres. The two most notable and crucial proteins are the sequence-specific binding proteins telomeric repeat binding factor 1 and 2 (TRF1 or TERF1 and TRF2 or TERF2). The latter one is best known for its contribution to T-loop formation and its essential role in telomere protection. But also TRF1 is an essential protein involved in the regulation of telomere length. Both proteins negatively regulate telomere length. Furthermore, other proteins are in interconnection with TRF1 and 2. The so-called POT1 (protection of telomeres 1 protein) is also a sequence-specific binding protein. Whereas TIN2 (TRF1-interacting nuclear factor 2), TPP1 (a POT1-binding partner) and Rap1 (repressor activator protein 1) are protein-specific binding proteins. The sum of all six proteins is termed shelterin complex [12-14] (**Figure 2**).
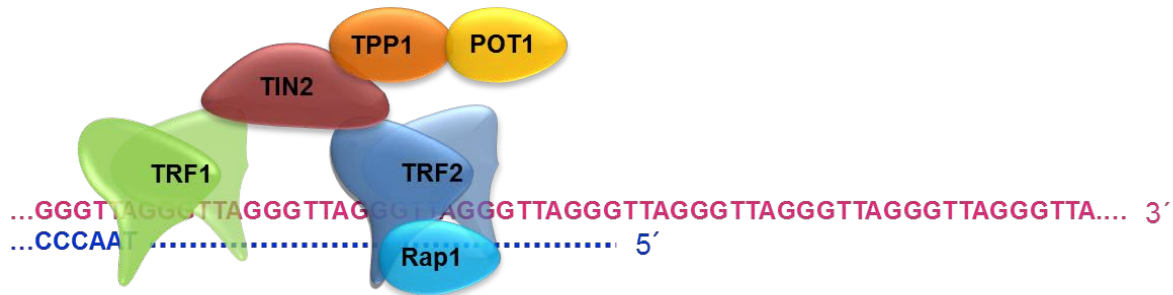
Figure 2: Shelterin Complex; TRF1 and 2: telomeric repeat binding factor 1 and 2, TIN2: TRF1-interacting nuclear factor 2, POT1: protection of telomeres 1 protein, TPP1: POT1-binding partner, Rap1: repressor activator protein 1. Figure adapted according to De Lange [12].

The T-loop formed by the shelterin complex at the end of telomeres prevents from being recognized as DNA double strand breaks and therefore from being inappropriate repaired by NHEJ (non-homologous end-joining) or HRR (homologous recombination repair) which would result in chromosome and genome instability [15]. But the role of telomeres is now known to be much more than protection. They are involved in the ageing process of cells, therefore they are considered to be a biological clock [10, 16].

With aging process, linear DNA of somatic cells shortens progressively at the 3'-end due to the inability of DNA polymerase to completely replicate to the very end as removal of the terminal RNA primer which is required at the lagging strand leaves a gap that cannot be refilled. Thus telomeres shorten with each cell division ('end-replication-problem' [17]). When telomere length has become critically short (Hayflick limit [18]), cellular senescence or apoptosis are mediated by the cell cycle checkpoint pathways, especially by the activation of tumor suppressor protein p53 and/or retinoblastoma protein pRb and cyclin-dependent kinase inhibitors p21 or p16 which results in cell cycle arrest (replicative senescence) [19-23]. If a cell escapes this cell cycle checkpoint and further replicates accompanied by increasing genome instability, there is a second checkpoint known as crisis. In rare cases cells overcome also the latter checkpoint and reach immortalization as also observed within various types of cancer [7, 23]. Additionally to the end-replication problem, telomeres of somatic cells shorten as a result of oxidative stress [24] and inflammation [25]. As telomeres consist to a high portion of the DNA base guanine, which is more likely to be oxidized as other DNA bases, oxidative stress is believed to be the major mechanism responsible for telomere shortening. The longer the telomeres the larger is the

point of attack by reactive oxidative species [26, 27]. The variation in telomere length is not only considered to be a result of environmental factors influencing the cellular pathophysiological conditions but is also ascribed to genetic influences. Large epidemiological genetic association studies showed a positive association of paternal age with telomere length [28-31]. Additionally a maternal as well as paternal pattern of inheritance has been suggested [30-35].

To overcome telomere shortening stem cells are able to express the enzyme telomerase, a specialized evolutionary conserved reverse transcriptase (encoded by TERT gene) that brings along its own RNA template (encoded by TERC gene) and synthesizes telomeric DNA to maintain telomere length [36-38] and the potential of self-renewal [39]. Telomerase is most active in germ line cells but also in progenitor cells, which are responsible for reparative processes in various tissues and therefore exposed to a high cell turnover. The view that telomerase is inactive and telomere length only reflects the expired number of cellular replication processes in differentiated somatic cells [40, 41] is under discussion. There is growing evidence that also telomere-lengthening is a biological relevant process [42-45]. The general population-based study by Huzen et al comprising 8,074 individuals reported after a median follow-up of 6.6 years a shortening of telomeres in 44% of subjects, stable length in 22% and an elongation in 34% of the subjects [44]. A similar observation was demonstrated recently by Weischer et al: of 4,576 individuals included from the general population with two RTL measurement time points 10 years apart, 56% of participants showed a loss of telomere length whereas 44% presented themselves with a telomere length gain [45].This phenomenon of telomere shortening and -lengthening is described as an oscillating dynamic process whose amplitude is decreasing with age [46]. This indicates that also other cell populations might have the ability of telomerase expression. A special mechanism of telomere-lengthening is described in 10-15% of tumor cells. This pathway is known as ALT (alternative lengthening of telomeres) pathway which elongates telomeres through homologous recombination [47-50] (**Figure 3**). Current results indicate that although a telomere shortening over life-time is the rule, a biological relevant and tumor-independent lengthening might be the truth (**Figure 4**). Therefore, the term biological clock as a synonym for telomere length will be more and more in the center of discussion.
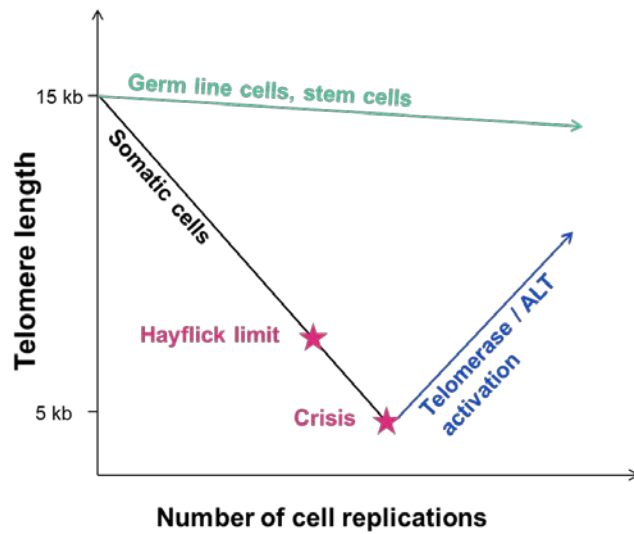
Figure 3: *Dynamics of telomere length: Cellular senescence, apoptosis, biological relevant lengthening of telomeres and cellular immortalization because of telomerase and/ or ALT (alternative lengthening of telomere). Figure adapted according to Royle et al [47].*
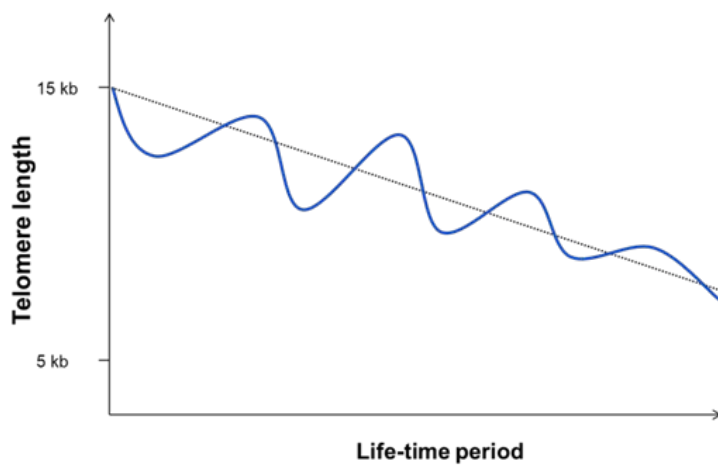
Figure 4: *Discussion of the "true" telomere dynamics: telomere shortening and -lengthening is described as an oscillating dynamic process whose amplitude is decreasing with age [46]. Figure adapted according to Svenson et al [46].*

2.3   MEANING OF TELOMERE LENGTH IN CHRONIC DISEASES

There is accumulating evidence that telomere dynamics represent a biological relevant process with an oscillating dynamic and amplitude decreasing with age [46]. Besides the relation to age, short telomere length is associated with several further disease risk factors in the general population such as male gender, obesity, hypertension, diabetes and smoking [45, 51, 52]. Telomere shortening is associated with an advanced burden of oxidative stress [24] and inflammation [24, 25]. Patients with chronic kidney disease, atherosclerosis and atherosclerosis-related diseases such as cardiovascular disease and peripheral arterial disease are not only exposed to a higher level of oxidative stress but also to systemic chronic inflammation accompanied by increased secretion of pro-inflammatory cytokines [53, 54]. These cytokines are proposed to induce a higher loss of telomeric repeats as result of a higher cell-turnover [55] and an enhanced activation of the respiratory chain which leads to production of reactive oxygen species [56] and furthermore to DNA damage and erosion [57]. Therefore, it is obvious that shortened telomere length has been shown to be associated with atherosclerosis [58-63], chronic kidney disease [64-66] and related cardiovascular disease [58, 67]. This observation coupled with the knowledge of variability of individual's telomere length at birth [68] has led to the assumption that reduced telomere length is an indicator of biological age and a potential marker and/or predictor of disease risk and disease progression [69, 70]. The potential causal relationship between telomere attrition and the pathogenesis of age-related diseases is not elucidated yet. But the strongest evidence for causality derives from a genome-wide association study comprising almost 50,000 individuals (**Figure 5**). The study identified genetic variants from seven genetic loci encoding proteins with known function in telomere biology to be associated with shorter telomere length. A risk score combining these seven lead variants was demonstrated to be associated with an increased risk for coronary artery disease in a study of 22,233 coronary artery disease patients and 64,762 control subjects [70].
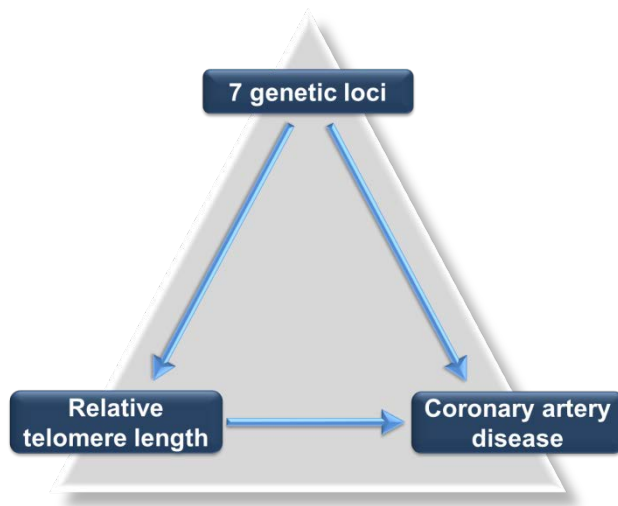
Figure 5: *Scheme of Mendelian randomization based on the genome-wide association study including almost 50,000 individuals by Codd et al [70] : genetic variants of 7 genetic loci encoding proteins with known function in telomere biology are associated with shorter RTL. A risk score of these 7 loci is associated with a higher risk of CAD. This finding supports causality between RTL and CAD. Figure adapted according to Kronenberg and Heid [71].*

Mendelian randomization [71-75] is a statistical approach applied on observational study data which uses measureable, functional-enlightened and well-understood genetic variations (in this case: 7 genetic loci) to assess the putative causal effect of a specific factor (in this case: relative telomere length, RTL) on a specific disease (in this case: coronary artery disease, CAD) without the need of an experimental study. The basis of this method is the biological fact that passing on mother's and father's allele variants of the 7 genetic loci to the child occurs randomly during conception. Since these genetic variants do not change during life and are not influenced by reverse causation or confounding, there is a stable even if small influence [71] on relative telomere length. Therefore, as has already been shown, RTL is associated with a higher risk for coronary artery disease [67, 76-78]. Also the risk score of these 7 loci is associated with a higher risk of CAD [70]. To summarize, although not clarified in detail, telomere length might not only be a marker of atherosclerotic diseases but might causally be involved in the pathogenesis of the disease when related to the cellular aging of the vascular system [70, 79, 80].

2.4   METHODOLOGICAL BACKGROUND

TELOMERE LENGTH MEASUREMENT

Several techniques of telomere length measurement have been published so far. The widely used procedures are telomere restriction fragments analysis by Southern blot [81, 82] and real-time quantitative polymerase chain reaction (qPCR) [83]. There are other methods such as quantitative fluorescence in situ hybridization (Q-FISH) [84, 85], Flow-FISH [86] and single telomere elongation length analysis (STELA) [87, 88]. As there are still controversies about the 'best' method to ascertain telomere length, the most common methods are described briefly in the following section. The aim is not to determine the 'best' method and not to announce the methodological details but to illustrate how data were generated in the presented projects and how data correlate between most common measurement methods.

Southern Blot

A quantitative approach to ascertain mean telomere length, the terminal restriction fragment analysis based on Southern blot hybridization using probes against telomeres was developed 20 years ago [81]. By degrading genomic DNA to short fragments with restriction enzymes, which leave intact the telomeres, followed by separation of telomeric DNA by gel electrophoresis and embedding on a nylon membrane a DNA smear is identified by hybridization of a telomeric probe for detection. This smear is the basis for the calculation of mean telomere length by comparing the smear to a known size marker or by densitometry (**Figure 6**).
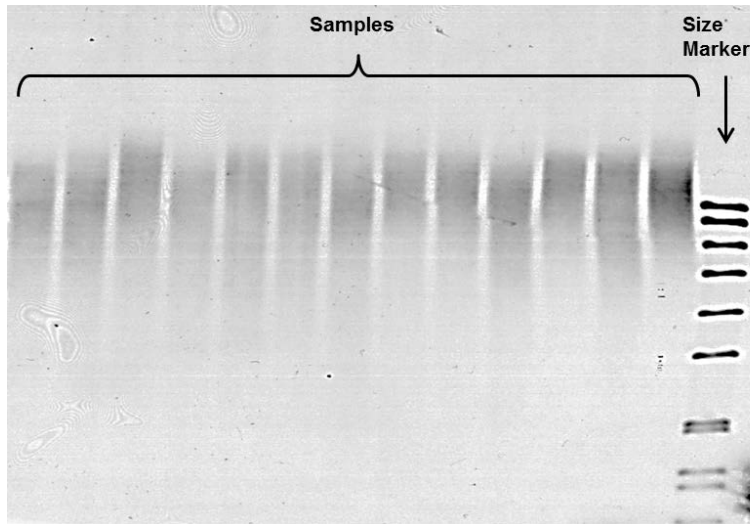
Figure 6: *Example of a Southern Blot to measure telomere length performed in our laboratory (sample description can be found in the section 2.4.1.3. Comparison of measurement methods). Separated DNA is presented as a DNA smear. Size marker is shown but in our case mean telomere length was calculated by densitometry.*

Real-time quantitative polymerase chain reaction (qPCR)

Relative telomere length measurement using a quantitative polymerase chain reaction (qPCR) approach was developed by Cawthon [83]. For our purposes the protocol was modified and extended to a highly standardized, as automated as possible, high-throughput approach with regard to control samples and data processing to ensure reliable and high quality data for further epidemiological studies. The premise of this method is to measure signals of a telomere (T) with specially designed primer sequences and a housekeeping single copy gene (S), which finally leads to T/S-ratio representing relative telomere length (RTL) (**Figure 7**). T/S-ratios are proportional to individual relative telomere length. In the presented projects the relative quantities were determined by the efficiency correction method [89]. This mathematical model calculates the ratio of a target gene (telomere) from the efficiencies and Ct-values (fractional number of cycles needed to reach the fluorescence threshold) of an experimental sample versus a reference gene (housekeeping gene) referring to a standard.

$$\text{Relative T/S} - \text{Ratio} = \frac{\text{eff (tel, sample)}^{\text{Ct (tel,sample)}}}{\text{eff (ref, sample)}^{\text{Ct (ref,sample)}}} \div \frac{\text{eff (tel, standard)}^{\text{Ct (tel,standard)}}}{\text{eff (ref, standard)}^{\text{Ct (ref,standard)}}}$$

Relative T/S-ratios reflect relative telomere length differences of the samples versus a positive control (standard DNA) and in comparison to the reference gene. In our projects the standard DNA sample was from a single individual.
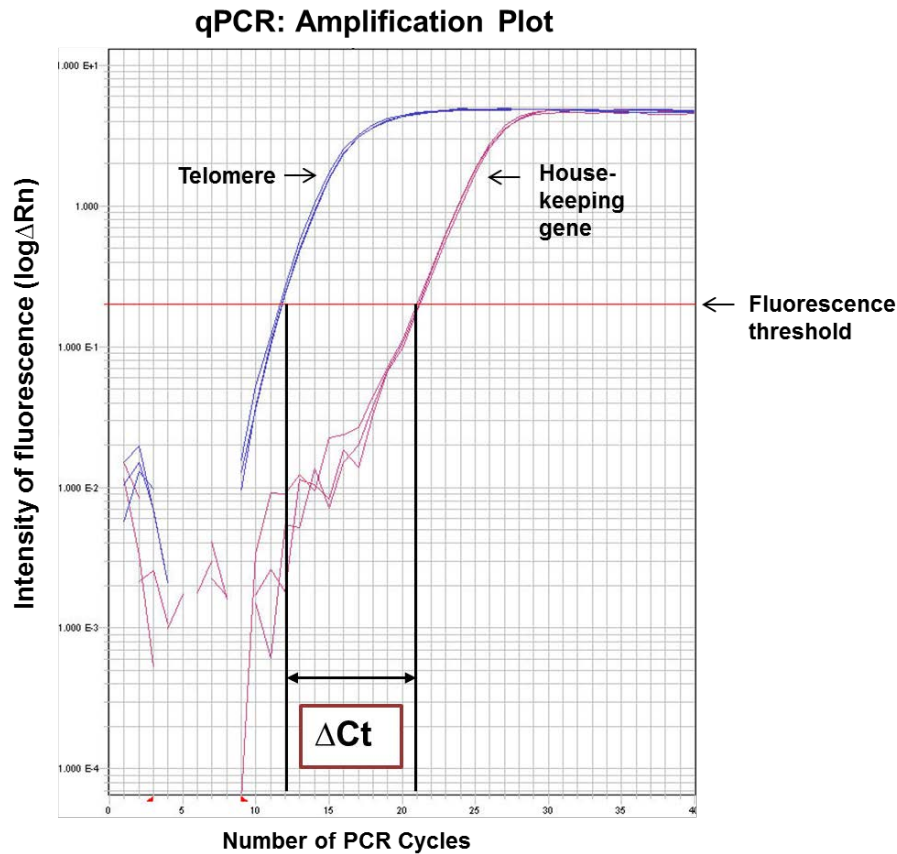

Figure 7: *Graphical example of raw data obtained by qPCR of one single subject (quadruplicates); x-axis: PCR cycle number; y-axis: log (ΔRn) indicates logarithmic Rn (=fluorescence intensity of the applied dye/fluorescence of passive reference dye) minus the baseline; blue line indicates amplification performance of telomeres, pink line indicates amplification performance of house-keeping gene; PCR efficiency can be evaluated from the slope of amplification plot. ΔCt-value (=Ct_{House-keeping gene} minus Ct_{Telomere}) and efficiency are the basis for the relative telomere length calculation. Equation was applied as described before.*

Comparison of measurement methods

There have been a lot of debates about the use of telomere restriction fragment analysis versus quantitative polymerase chain reaction (qPCR) [83]. Arguments pro qPCR include the experience that the telomere restriction fragment technique is labor-intensive,

difficult to quantify, requires a minimum of 3µg of DNA [82] and the fact that high variable length of subtelomeric DNA confounds the result, while qPCR requires far less DNA, measures only the length of pure telomere sequence and is a high-throughput method even if telomere length is only quantified relatively.

To compare measurements of telomere length obtained with qPCR with conventional telomere restriction fragment analysis, we typed 38 DNA samples obtained from donors from Central Institute of Blood Transfusion and Immunology, University Hospital, Innsbruck, Austria by using qPCR as well as telomere restriction fragment analyses. Correlation results showed a Spearman correlation coefficient r of 0.83 (p<0.0001) (**Figure 8**). Similar results were demonstrated by others [58, 67, 83].
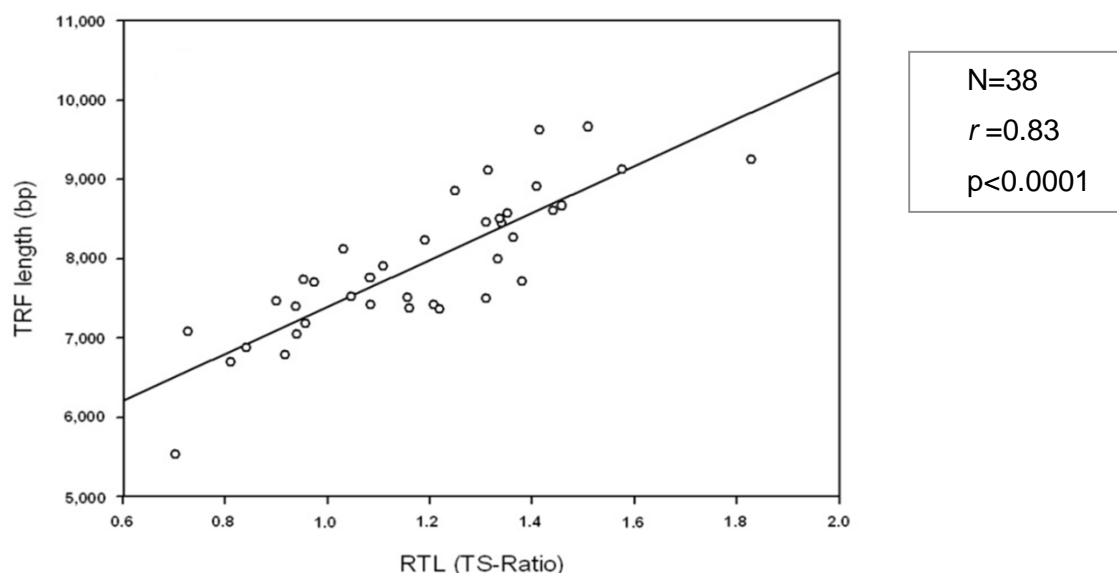

Figure 8: Correlation between relative telomere length (RTL; expressed as TS-Ratio) and absolute telomere length (telomere restriction fragment length; expressed as TRF length) of the 38 typed DNA samples obtained from donors from Central Institute of Blood Transfusion and Immunology, University Hospital, Innsbruck, Austria

The largest study to date examining the comparability of qPCR and Southern blot was published recently by Elbers et al [90]. Results in 681 participants showed that the correlation between the data generated by qPCR and Southern blot are only modest (r =0.52). One explanation might be an interdonor variability in length of subtelomeric DNA, which is measured by Southern Blot method but not with qPCR. Moreover, different measurement error has to be considered. Nevertheless, the age and gender effect -

usually men are reported to have shorter telomere length - was detected by both methods [90]. This might indicate that the strength of the relationship between telomere length and age as well as telomere length and sex is much stronger as reported to date.

## ASSOCIATION ANALYSIS AND STUDY DESIGN

Association studies are one of the central methods in genetic epidemiology to evaluate the interrelation of a parameter of interest such as telomeres, genes or gene phenotypes with various phenotypical parameters or diseases (**Figure 9**). This study type may possibly indicate causality but neither the direction of the effect nor a final causal conclusion can be determined.
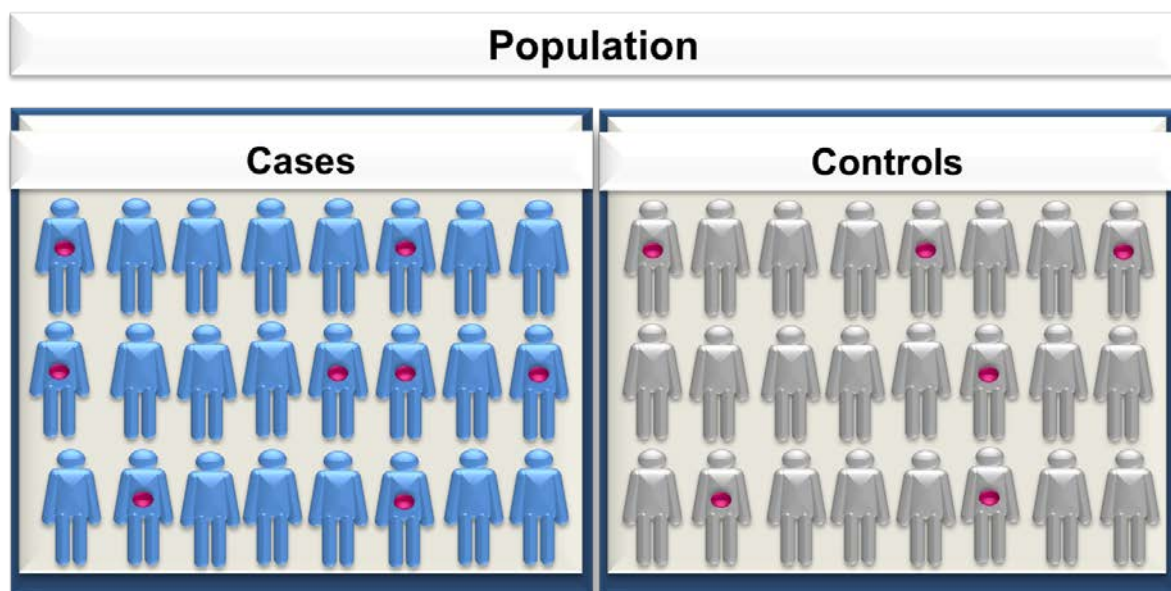
Figure 9: Scheme of an association study. Cases (blue figures): individuals with a defined disease. Controls (grey figures): healthy individuals, red dots: specific parameter of interest (e.g. genetic variant) is present. In the group of cases 25% more subjects are carrier of the parameter of interest as in controls. Figure adapted according to Hirschhorn and Daly [91].

An appropriate study design is of major importance for association studies. This type of epidemiological studies can be classified in two groups: intervention studies and observational studies. Observational studies are comprised of three central types: case-control studies, cross-sectional studies and prospective cohort studies. Case-control

studies and cross-sectional studies are often used for hypotheses generation and thus can serve as starting point for subsequent prospective cohort studies.
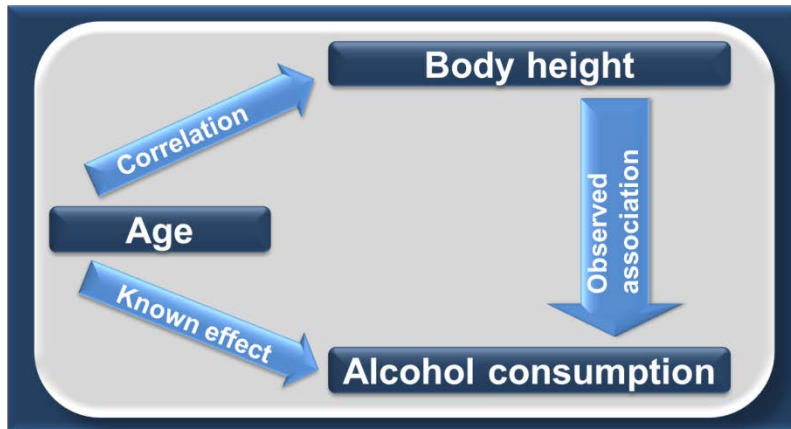

Figure 10: Example of confounding in a group of individuals between 5 and 20 years of age: age is the confounder in the association of body height and alcohol consumption.

A major problem of observational studies is confounding (**Figure 10**). This implicates that the recruited subjects do not only differ in the investigated main parameter but also in other characteristics which are often difficult to quantify; for example socioeconomic status or physical constitution. Confounding is a key element towards evaluation of associations between telomere length and diseases as it has been shown that telomere length is related to many risk factors of various multifactorial diseases. In general, genetic association studies are less influenced by confounding effects as genetic markers are randomly distributed throughout populations. However, the possibility to control for potential confounding effects is a major advantage of genetic association studies. Controlling can be done by applying restriction, matching, stratification or adjustment. The first step of controlling can be done by restriction, e.g. by selection of only men or only smokers. Second option of confounder-controlling is matching, e.g. matching cases and controls pairwise by the confounder. This method enables to deactivate the influence of individual confounders. But there is also the risk to overmatch followed by misrepresentation of the population. Furthermore, stratification can be of valuable assistance and is an often adopted approach. It allows the analysis of an effect in all single strata of the confounder variable and allows thereby investigating the association without an effect of the confounding variable (e.g. GFR strata as can be found in

3.2 Project 2: Relative Telomere Length and Cardiovascular Disease in a Chronic Kidney Disease Cohort). In addition, adjustment for the confounder in multivariate models which might influence results relevantly provides an opportunity of controlling for confounding.

For further details on epidemiology and study design see "Epidemiologie" by Leon Gordis [92].
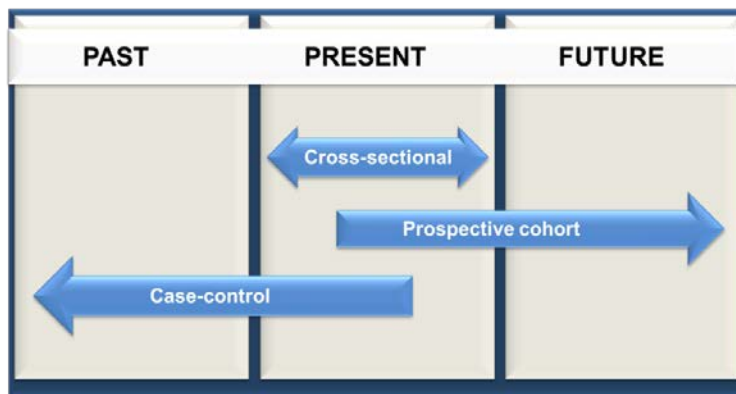
Figure 11: *Three central types of study design including time of enrollment. All studies are observational studies. Cross-sectional study: snapshot of a population at one time point; Prospective cohort study: population is followed over a time period and incidence of outcome is investigated, Case-control study: retrospectively comparing patients who have already been diagnosed with the disease (=cases) to controls. Figure adapted according to Levin [93].*

Case-control study

The case-control approach is the simplest design to evaluate associations. According to a clear definition of a certain disease a homogenous group of cases is selected at the study start point. In contrast the control group either comprises healthy study subjects or subjects recruited randomly from the population without the disease of interest but possible other unknown diseases specified by predefined inclusion- or exclusion criteria. To assess associations, cases and controls are compared. Typically the case-control study design is applied for investigations of rare diseases or diseases with a long period of disease manifestation (e.g. peripheral arterial disease). The design allows recruiting individuals screened already positive for presence of disease and therefore the data collection can be conducted within a short period of time. In addition much less study

participants are required compared to cohort studies. But the vulnerability especially concerning selection bias (e.g. composition of the appropriate control group) as well as survival bias (e.g. investigating diseases with a high risk for a fatal outcome, such recruited individuals might not be representative) is a major drawback of case-control studies. Within the PhD thesis this study design was applied for the CArdioVAScular disease in patients with Intermittent Claudication (CAVASIC) Study [59], which comprises male patients with intermittent claudication and age- and diabetes-matched volunteers as control group.

Cross-sectional study

A cross-sectional study is conducted at one single time point and provides a snap-shot of associations between genotypic/phenotypic parameters and the outcome of interest. Study subjects represent a previously defined population (e.g. Caucasian individuals within a defined age range and living in a specific geographical region). This type of studies is a useful tool to estimate the prevalence of common diseases (e.g. cardiovascular disease) or to determine the distribution of potential risk factors (e.g. alcohol consumption, smoking). A further advantage is the possibility to study multiple outcomes. In addition, it can be a useful and relatively inexpensive − in contrast to prospective cohort studies − approach to evaluate new hypotheses, either for functional studies or for prospective cohort studies which both require large financial resources.

Prospective cohort study

In a prospective cohort study the cohort is recruited before the investigated outcome (e.g. a disease) becomes manifest. There are two alternative strategies of study population selection: first, individuals who are exposed or not-exposed to certain risk factors or diseases (e.g. population with chronic kidney disease). Second option is recruitment based on the entire population without selection of a certain exposition-dependent factor (e.g. population from the same geographical region). This ensures that the cohort is comprised by individuals sharing a common characteristic (applies for both recruitment strategies). A typical example of a population-based study is the Bruneck Study [94]. This study is conducted to evaluate atherosclerosis, its related diseases and risk factors and is performed in Bruneck, a small city in South Tyrol (northern Italy) comprising about 15,500 inhabitants. The study population was selected as a random sample of the

resident individuals aged between 40-79 years by regarding a balanced gender distribution. In addition, a typical strategy for stratifying was applied – recruitment of a specific number of subjects for each age decade. In contrast to the Bruneck Study, the German Chronic Kidney Disease (GCKD) Study [95] invited already diseased patients – a typical example of a disease-based study. Regardless of type, study subjects are followed over a defined time period with the aim to monitor incident outcomes (e.g. for Bruneck Study disease occurrence or for GCKD Study disease progression). Additionally, data of parameters which are thought be related to the outcome of interest are collected. Main causes of bias are selection bias and loss of follow-up.