Association between the number of CAG repeats in polymerase gamma and Parkinson disease in the Norwegian population

Novin Balafkan

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Science in Medical Cell Biology

Department of Biomedicine / Department of Clinical Medicine

Faculty of Medicine and Dentistry

University of Bergen

Spring 2012

Acknowledgements

First and foremost I would like to thank of my supervisor Professor Laurence Bindoff who has supported me throughout my thesis with his abundantly helpful and invaluable guidance, his patience and knowledge. This research project would not have been possible without my co-supervisor Dr. Charalampos Tzoulis who introduced me to the project and patiently thought me about the methods and graciously lending his time and expertise in raising my knowledge. I attribute the level of my Master degree to his encouragement and effort and without him this thesis would not have been completed or written. I also need to thank lab members for all their helps during this time.

Finally, I want to convey thanks to the Faculty of Medicine for providing the financial means and laboratory facilities during these two years.

Novin Balafkan

May 2012

Table of contents **Page**

List of abbreviations

EtBr	ethidium bromide
GWAS	genome wide association studies
HD	Huntington's disease
kDa	kilodalton
kb	kilobase
LD	linkage disequilibrium
mtDNA	mitochondrial DNA
n.a.	not analyzed
OXPHOS	oxidative phosphorylation
PEO	progressive external ophthalmoplegia
Polγ	polymerase gamma protein, catalytic subunit
PD	Parkinson disease
POLG1	polymerase gamma gene, catalytic subunit
Poly Q	polyglutamine
PCR	polymerase chain reaction
ROS	reactive oxygen species
STR	short tandem repeats
TM	melting temperature
TNRs	trinucleotide repeats
TREDs	triplet repeat expansion diseases

Summary

Parkinson disease (PD) is a common neurodegenerative movement disorder that increases in prevalence with age. It is characterized clinically by resting tremor, rigidity, bradykinesia, and failing balance, and is due to the gradual loss of the dopaminergic neurons in the substantia nigra of the midbrain. Different factors including environmental (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), 6-hydroxydopamine (6-OHDA), etc.) and genetic (PINK1, DJ1 and LRRK2) have been found to cause the loss of the dopaminergic neurons in the substantia nigra. In addition, deficiency of respiratory chain complexes (complex I, II and IV) and increased levels of mtDNA deletions have been identified in PD patients. All of these factors, from the environmental to the genetic and the biochemical converge on one site, namely the mitochondrion: MPTP poisons complex I of the respiratory chain, PINK1, DJ1 and LRRK2 have all been linked to mitochondrial and lastly, mutations in the catalytic subunit of the mitochondrial DNA polymerase, *POLG1*, are known to cause Parkinsonism.

The *POLG1* gene also contains a trinucleotide (CAG) repeat region and while the majority of individuals have 10 CAG repeats (10Q) this region is also unstable and both longer and shorter repeats are seen. Previous studies have highlighted an association between CAG length in *POLG1* and the development of PD. This has been done in different population such as England and United states, also for Nordic population (Finland and Sweden). In most populations studied, an association has been found, but in one population (England) no association was seen. The aim of this study was, therefore, to ascertain whether the length polymorphism of the CAG tract in *POLG1* was associated with PD in the Norwegian population. In order to do this we studied cohort of Norwegian patients and age-matched controls that had been collected and clinically characterized as part of the ParkVest study.

We found a significant association between the non 10Q alleles (P= 0.027; OR 1.53), especially 12Q (P= 0.031; OR 2.38), and PD. In addition, meta-analysis of all available studies confirmed a highly significant association between the non 10Q alleles (P= 0.0004; OR 1.27) particularly 12Q (P= 0.0047; OR 1.55) and PD. This pooled analysis also showed significant association between 9Q and PD (P= 0.0427; OR 1.21). Our results suggest that CAG repeat polymorphism in the *POLG1* can be a risk factor or marker for PD.

1. Introduction

1.1 Mitochondria

1.1.1 History and Origin

Mitochondria originated between 3.45 and 2 billion years ago from an endosymbiotic relationship between an α-proteobacterial organism and an anaerobic host cell. Mitochondria are known as the "power houses" of the cell due to their ability to produce energy in the form of adenosine triphosphate (ATP) from carbohydrates and fat by the oxidative phosphorylation (OXPHOS) process. Aerobic bacteria would have provided an aerobic cell which can carry out oxidative metabolism by proteins that encode in α-proteobacterial genome (1).

Rickettsia prowazekii, an intracellular parasite, has a similar genome to mtDNA and is possibly related to the mitochondrial ancestor. Most of the original mitochondrial genes have been translocated to the nucleus during evolution and today only 13 peptide genes are still encoded in mtDNA (2, 3).

1.1.2 Structure

Mitochondria appear as round or oval-shaped structures under the electron-microscope (figure 1). In reality, they form a dynamic network or reticulum undergoing continuous fusion and fission (4). Mitochondria are bilayer organelles consisting of an inner space called matrix space, an outer membrane (OM) and an inner membrane (IM) with a space in between that is called the intermembrane space (IMS) and the space surrounded by the inner membrane is the matrix which contains a high number of enzymes (proteins), RNAs and mitochondrial DNA and is segmented by segments that extend from the inner membrane called cristae (4).

The outer membrane plays a role as the barrier and encloses the mitochondrion. It has porins and non-selective membrane channels formed by integral membrane proteins (5). Unlike the outer membrane there is no porin in the inner membrane and it is a highly impermeable layer. Ions and molecules are transported through this layer by the help of different *transporters* (carrier molecules). Proteins responsible for the redox reactions of oxidative phosphorylation and ATP synthase (the respiratory chain) that generates ATP in the matrix are present in this inner membrane. The mitochondrial inner membrane contains a high amount of

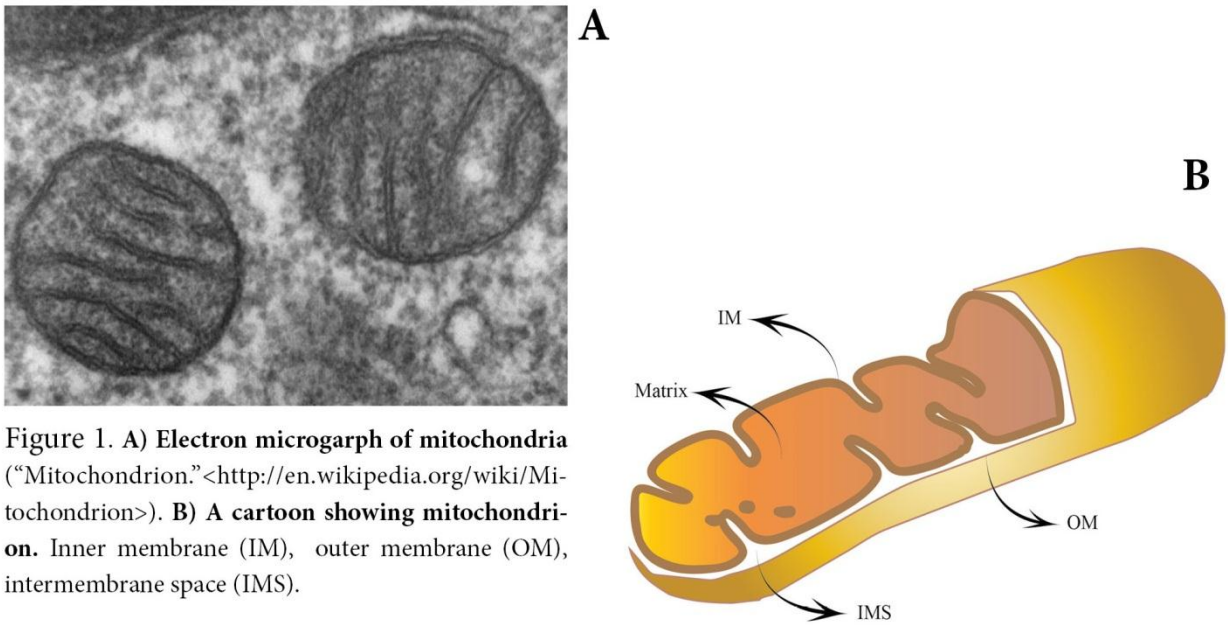


Figure 1. **A) Electron microgarph of mitochondria** ("Mitochondrion."<http://en.wikipedia.org/wiki/Mitochondrion>). **B) A cartoon showing mitochondrion.** Inner membrane (IM), outer membrane (OM), intermembrane space (IMS).

protein (75 %) compared to other membrane-bound organelles (usually 50 % protein and 50 % lipid) (1). The amount of inner mitochondrial membrane per cell depends on the bioenergetic parameters (energy requirements of the cell) in the cell which can change the size of mitochondria instead of increasing the number of mitochondria in the cell (6, 7). The number of mitochondria in each cell is variable and can reach into several thousand in high energy demanding tissues like neuronal tissue, cardiac muscle etc., and the number of mtDNA in each mitochondrion varies between 2-10 (8).

1.1.3 Function

Mitochondria are involved in variety functions including fatty acid metabolism, urea cycle, heme biosynthesis and lipid biosynthesis and metabolism (9). The focus of this thesis is on one of the most important functions of mitochondria which is converting the energy in different metabolic fuels such as glucose, to the chemical bound energy in ATP molecules which supplies of the energy in most biological reactions.

Glucose is one of the major sources of energy and is broken down by glycolysis into pyruvate. Glycolysis occurs in the cytosol and its end product, pyruvate, must be transported into mitochondria where it is converted to acetyl CoA that then enters the citric acid cycle (Krebs cycle). Acetyl coA is further metabolized by the Krebs cycle to produce carbon dioxide and water and during this process, cofactors are reduced (NAD^+ to NADH and FAD to $FADH_2$).

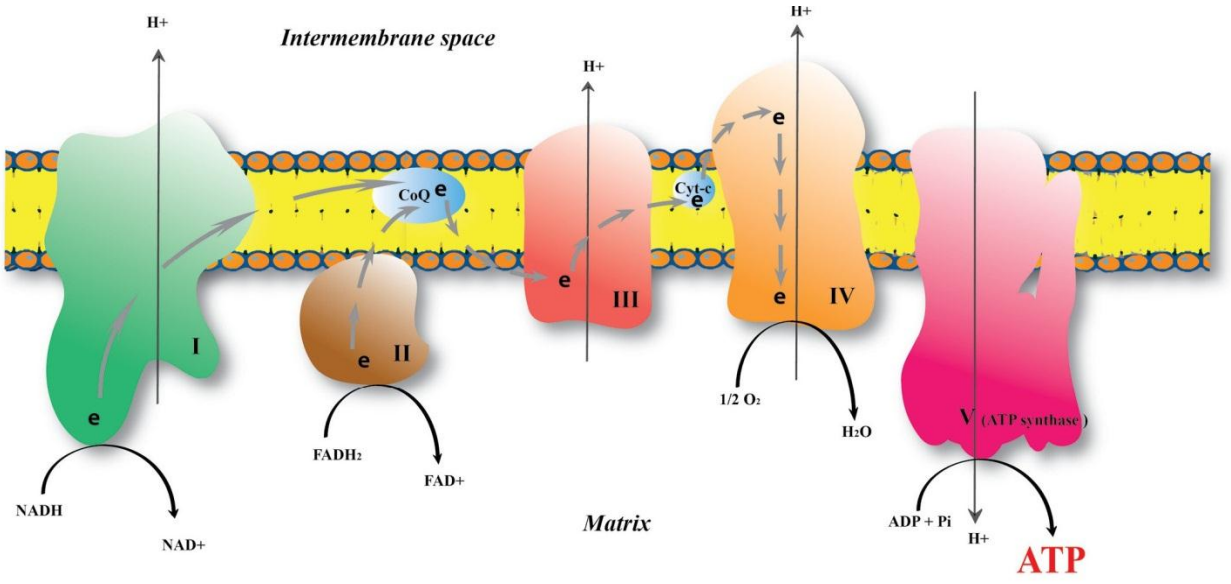


Figure 2. **Overview of the respiratory chain structure.**

NADH and $FADH_2$ are electron carriers and during oxidative phosphorylation, which takes place in the inner membrane of mitochondria, electrons are donated from NADH and $FADH_2$ to either complex I or II of the respiratory chain. The process of electron transport from one complex to another is associated with the release of energy and synthesis of ATP from ADP (1). The electron transport chain consists a number of different protein complexes (figure 2), Complex I (NADH: ubiquinone oxidoreductase), Complex II or succinate dehydrogenase, Complex III or cytochrome bc_1 complex, Complex IV and Complex V. Complex V or H^+–ATP synthase is the last complex in respiratory chain which is activated by H^+ flow from innermembrane into the matrix resulting in ATP synthesis from ADP and Pi.

1.1.4 Genome

Mitochondria have their own 16.5 (kb) circular double-stranded DNA genome which is composed of 37 genes comprising 13 peptides, 22 tRNAs and 2 rRNAs (figure 3). Mitochondrial DNA does not have any introns, but there is a non-coding region which is known as the "D-loop" and is the site for the origin of replication and the promoters for transcription. The peptides which are encoded by mtDNA are subunits of the respiratory chain complexes I (ND1-ND6, ND4L), III (CYB), IV (COX I-IV) and V (ATPase 6, ATPase 8). Complex II is encoded exclusively in the nucleus. All other proteins such as the other components of the respiratory

chain, different enzymes, chaperones (protein) and replication machinery, including Pol γ (DNA polymerase gamma) are encoded by nuclear genome (10, 11).

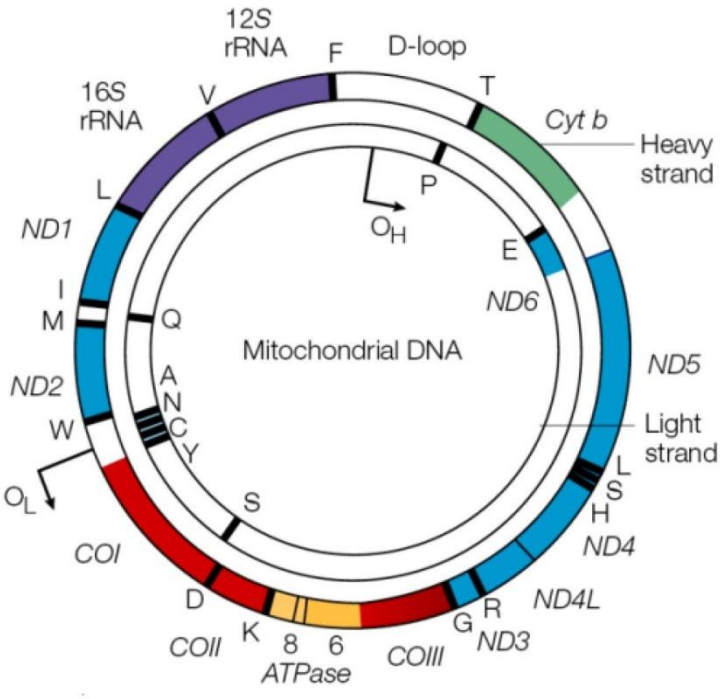


Figure 3. **A cartoon showing mitochondrial DNA structure.** MtDNA encodes 37 genes, 13 proteins and 24 structural RNA. ND1-ND6 and ND4L are porteins involved in complex I (blue); COI, COII and COIII encode subunits of Complex IV (Red); Cytochrome b is the only complex III protein encoded by mtDNA (green) and subunits of the ATP synthase (ATPase 6 and 8) are shown in yellow. tRNAs (black lines and denoted by their single letter code) and rRNAs (purple) are required for mitochondrial protein synthesis. (12)

1.1.5 Replication and inheritance

Mitochondria are inherited exclusively from the mother, which makes it possible to use mtDNA in order to trace the maternal ancestry of populations (13). There are two models for mtDNA replication, the asynchronous strand displacement model and the strand coupled model. In asynchronous strand displacement model, replication of the heavy strand starts prior to replication of the light strand and there is a gap between replication start times of these two strands. In strand couple model both strand of mtDNA replicate simultaneously from replication zone. Mitochondria in human cells use both model to replicate its own DNA (10).

1.1.6 **DNA polymerase gamma** (DNA polymerase γ)

Polymerase gamma (Pol γ) is a DNA dependent, DNA polymerase consisting of a catalytic subunit (140 kDa) and accessory subunit (55 kDa)Human Pol γ is in fact a heterotrimer that is composed of one catalytic subunit that binds two accessory subunits. Accessory subunit functions include -increasing DNA binding affinity and primer recognition which is important in

the initiation of replication and increasing the holoenzyme structural integrity (14, 15). Exonuclease and DNA polymerase activity are the most important functions of Pol γ and these functions reside in the catalytic subunit. The accessory subunit is encoded by the *POLG*2 gene on chromosome 17q23-24 and catalytic subunit by *POLG1* gene on chromosome 15q25. The catalytic subunit consists of an exonuclease domain, a polymerase domain, and a linker region which is involved in DNA binding and interaction with the accessory subunits (figure 4) (16, 17). The Pol γ pre-protein has 1239 amino acids which is encoded by 23 exon and the mitochondrial signal is at the N-terminal of the *POLG* pre-protein which is cleaved off after translocation to the mitochondria (18). Pol γ repair function is mostly limited to base excision repair (BER) which use two pathways; single-nucleotide-BER (SN-BER) or long-patch BER (LP-BER) (19).

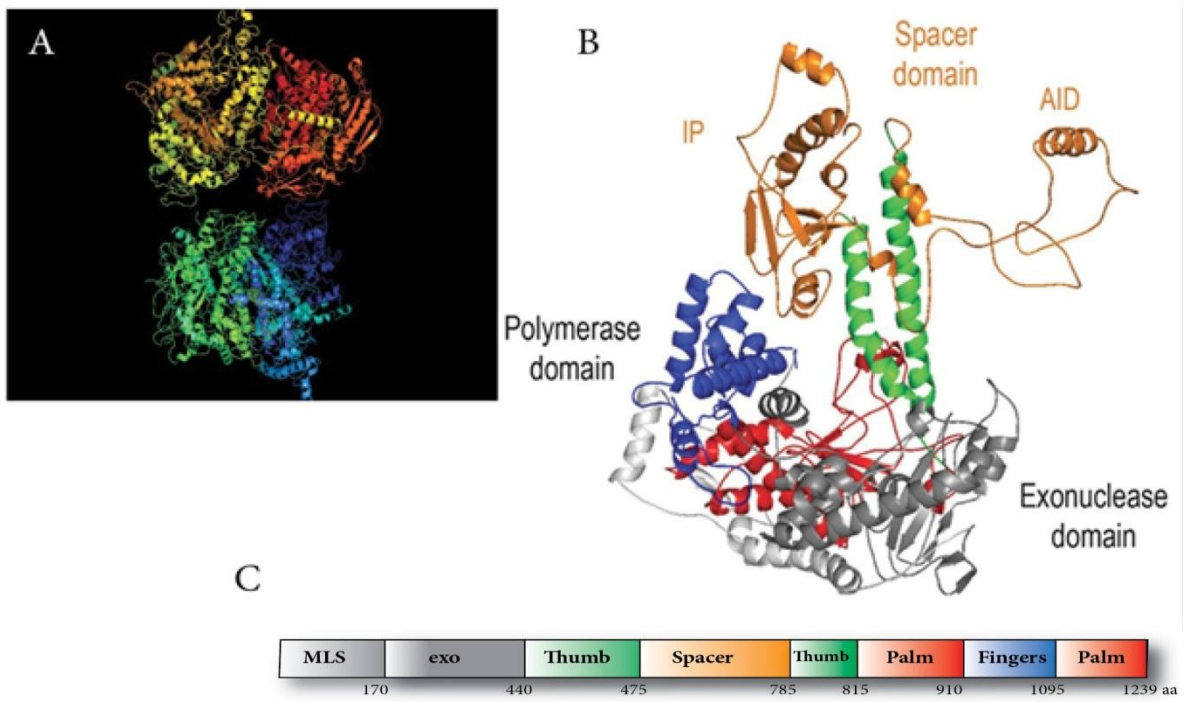
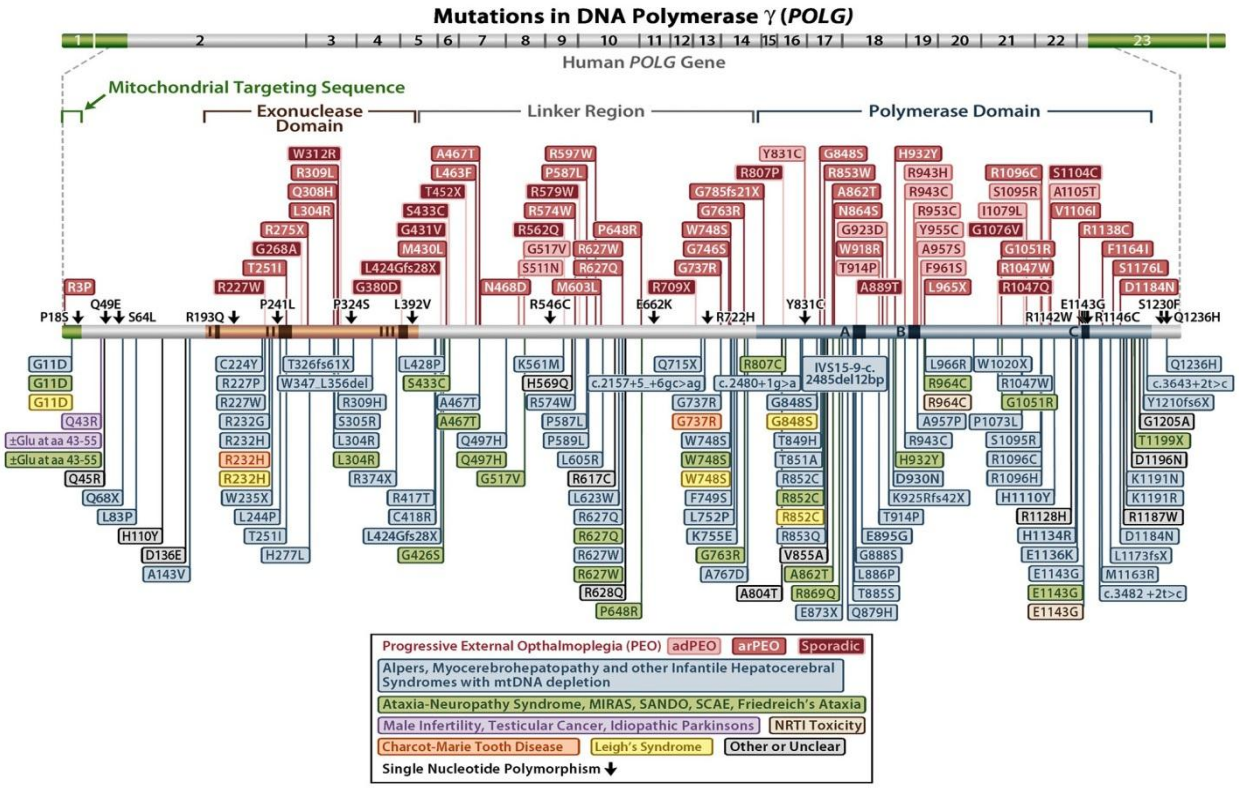


Figure 4. **Structure of Pol γ. A) Ribbon diagrams of the tertiary structure of Pol γ holoenzyme** (made with Pymol program). **B) Structure of Pol γ** A. Thumb (green), palm (red), and fingers (blue) subdomains and the exonuclease domain (gray). The spacer domain (orange) divided into two subdomains. **C) Domains of Pol γ A are shown in a linear form** (Y. S. Lee, Kennedy, & Yin, 2009).

1.1.7 **Human Pol γ mutation**

POLG1 point mutations are associated with a wide spectrum of disease (20).Point mutations in the *POLG1* gene can eliminate the polymerase or proofreading activity and lead to secondary defects in mtDNA during replication. This in turn, can cause of dysfunction of the respiratory chain. Almost 150 mutations in *POLG1* are listed by The Human Pol γ Mutation Database (http://tools.niehs.nih.gov/*POLG*/). Mutations in *POLG1* are spread through all domains (figure 5) and most of these mutations were found in patients with mitochondrial disease. Different mutations in *POLG1* allele in patients with mitochondrial disease can take place, which show strong evidence of contribution *POLG1* mutations to brain and liver damage in infants (Alpers syndrome) and neuromuscular tissues in adults (PEO) (19).

Figure 5. **Schematic diagram of human pol γ gene (top) and protein (bottom).**
Shows the location of amino acid substitutions resulting from mutations associated with disease (in the boxes) and neutral polymorphisms found in unaffected populations (arrows). (http://tools.niehs.nih.gov/POLG/).



1.1.8 Dysfunction and disease in associated with *POLG1*

POLG1 mutations are a common cause of mitochondrial disease and more than 150 pathogenic *POLG1* mutations have been reported (21). Autosomal dominant progressive external ophthalmoplegia (PEO) was the first disorder to be described, which was caused by mutations in *POLG1* (22).*POLG1* mutations cause a variety of clinical syndromes including autosomal recessive and dominant PEO (23), Alpers-Huttenlocher syndrome (24) and mitochondrial spinocerebellar ataxia and epilepsy (MSCAE) (25).

1.2 Parkinson's disease

Parkinson disease (PD) was first described in 1817 by James Parkinson in his paper "An essay on shaking palsy" (26). PD is a relatively common neurodegenerative disorder with an age depended prevalence (108 to 257/100,000 in the general population; 1280 to 1500/100,000 in the over 60 years of age). The incidence of PD in Norwegian population is 12.6 per 100,000 individuals and male-to-female ratio is 1.58 (Clinical onset of PD in women is 68.6 and 66.3 years for men) (27). PD is characterized clinically by resting tremor, rigidity, bradykinesia, failing balance and a variety of non-motor symptoms including cognitive and behavioral changes and sleep disorder. The course is slowly progressive leading to increasing motor disability and shortened life expectancy. The pathophysiology of PD involves gradual loss of the dopaminergic neurons in the substantia nigra of the midbrain, resulting in decreased nigrostriatal input and ultimately decreased cortical motor activity resulting in the hypokinetic character of the disorder (28).

1.2.1 Etiology of PD

1.2.1.1 Environmental factors

Several environmental factors are associated with PD. The compound methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) which is an inhibitor of the mitochondrial respiratory complex I selectively damages the dopaminergic neurons of the substantia nigra, presumably because it is specifically taken up by the dopamine reuptake receptors of the nigrostriatal terminals, and subsequently concentrated in mitochondria where it poisons complex I. 6-hydroxydopamine (6-

OHDA) is another neurotoxin that induces oxidative stress and generates active oxygen species in the form of superoxide (O2). It is conjectured that the increase ROS production level in the mitochondria leads to mitochondrial alterations and induces apoptosis in dopaminergic neurons (29, 30). Interestingly, smoking cigarettes and drinking coffee are protective factors in PD (31).

1.2.1.2 Genetic factors

Most PD cases are sporadic and idiopathic, meaning that the etiology is unknown. However, in recent years several genes have been identified causing familial PD.

1.2.1.2.1 Monogenic (Mendelian, familial) PD

Recently, mutations in genes such as α-synuclein, Parkin, *PINK1*, *DJ1* and *LRRK2* have been identified causing PD (32, 33). Recent studies show a significant association between mutations in *Parkin* and α-synuclein and cellular ubiquitin proteasomal system (UPS) which leads to decreasing clearance of unwanted proteins (34).

1.2.1.2.2 Genes in sporadic PD

While the great majority of PD cases are sporadic (non-familial), and only few caused by single gene defects, there is increasing evidence that genetic risk factors play a major role in the pathogenesis of sporadic PD. Genetic association studies comparing large groups of PD patients with healthy controls have shown mutant glucocerebrosidase can be a risk factor for the development of PD (35). Moreover, haplotype analysis shows α-synuclein gene polymorphism is a risk factor for the development of parkinsonism (36). One problem with genetic association studies is that they do not always directly identify risk variants. These studies do not necessarily identify the genes which are associated with PD, but can be interpreted as markers for the disease base on linkage disequilibrium (LD). Thus, these markers might be near neighbors to the disease allele which carries a mutation gene (*cis* with marker) or a disease allele in adjacent gene (31).

1.2.2 PD and Mitochondrial dysfunction

There is increasing evidence suggesting that mitochondrial dysfunction plays a role in the etiology of PD (37-39). Deficiency of complex I was shown in post-mortem brains of PD patients (40, 41). Deficiency of the mitochondrial respiratory chain complexes I, II and IV were also demonstrated in skeletal muscle samples from living patients with Parkinson's disease (42,

43). Other studies have shown more deletions in mtDNA in the brains of patients with PD compared to an age-matched control group (37). In addition, both PINK1 and Parkin are linked to mitochondria dynamics and function and it has been suggested that the association between mutations in these genes and PD is an increased oxidative stress and free radical damage (44, 45).

1.3 Trinucleotide repeats

1.3.1 Biology of trinucleotide repeats

Short tandem repeats (STR) of nucleotides such as mono-, di-, tri- or tetranucleotide repeats are usually classified as microsatellites. Even penta- or hexa- nucleotide repeats can be classified in this way (46). There is no minimum number of repeats or number of nucleotides for a tandem repeat to be referred to as a microsatellite. Almost 3% of genome comprises microsatellites meaning that there are more than one million microsatellite regions in the human genome. Moreover, the number of microsatellites increases at the ends of chromosomes. It seems there is a positive correlation between the size of genome and number of microsatellites and trinucleotide repeats have the lowest frequency compared to the other microsatellites (mono-, di-, tri- or tetranucleotide repeats) (46).

One of the intrinsic characters of microsatellites is their high length mutability and this feature makes them a good genetic marker in different type of studies such as population genetics, genetic mapping and linkage analysis (47). Trinucleotide repeats (TNRs) are one the most important class of microsatellites, due to their known role in triplet repeat expansion diseases (TREDs).

1.3.2 Trinucleotide repeat disorder

Instability in TNRs is the cause of several developmental and neuromuscular disorders. Mismatch repair deficiency is one of the reasons for expanding triplet repeats (48) Instability of TNRs is reported in different human disorders, including Huntington's disease (HD), myotonic dystrophy (DM), and several forms of spinocerebellar ataxia (eg SCA1) (49, 50). Nasir .et al. 1996 showed that expansion of the polyglutamine tract in huntingtin protein can lead to a

conformational change causing altered protein mobility which is considerable issues for forming neural network during embryonic development and it causations of Huntington disease (51).

1.3.3 trinucleotide repeats in *POLG1*

The human *POLG1* gene contains a unique feature that is absent in the mouse, drosophila or yeast: the human *POLG1* gene has a trinucleotide (CAG) repeat in exon 2 in the NH2-terminal domain of the gene. According to the NCBI Reference sequence (NG_008218.1) it contains 10 glutamine-encoding CAG codons, followed by a single CAA and two further CAGs (52).

Analysis of CAG trinucleotide repeats in *POLG1* in genomic DNA from single-cell clones (such as LS180, LS174T, AN3CA, etc.) showed this region of *POLG1* is unstable. Finding of different *POLG1* polymorphisms in one single-cell clone indicate expansion or contraction of the CAG repeat sequence. Also, the presence or intensity of each allele was varied from clone to clone. However in some clones such as SKUT-1 and DU145 the stability of CAG trinucleotide repeats region is reported (53).

Interestingly, recombinant *POLG1* with ΔQ_{10} (10 CAG repeat deleted) was completely stable in human cells and there was no changes in proofreading activity and mitochondrial function (oxygen consumption), however polymerase activity and Pol γ expression slightly increased and mtDNA content slightly decreased in compare to normal cell line in Spelbrink et al. 2000 study (18)

1.3.4 *POLG1* trinucleotide repeats and diseases

The investigation of unstable TNRs region in exonic part of *POLG1* was surprising because of the prominent function and sensitivity of this enzyme in mitochondrial biogenesis. Several studies tried to find about the possible association between instability of this trinucleotide repeats and different disease. Different studies suggest an association between the length of the CAG repeat in *POLG1* gene and breast cancer (54), testicular cancer (55, 56) and male infertility (57). The findings are inconsistent however, as other studies do not confirm these associations (58-62).

Different studies from Sweden (63), Finland (64) and United states (65) reported an association between the length alteration of CAG trinucleotide repeats and PD. These studies reported

significant association between non 10 repeats of CAG trinucleotide repeats and PD. However three different studies for England population (66-68) do not confirm these associations.

2. Aims of the study

Given the mounting evidence that mitochondrial dysfunction plays a role in PD together with early studies showing an association between non10/11Q tract lengths in other populations, the aim of this thesis was to investigate the possible association between the number of CAG repeats in *POLG1* gene and Parkinson disease in the Norwegian population. The sub aims were:

1. To investigate if the CAG repeat length was associated in idiopathic PD in Norwegian patients and if so whether this was specific for one of more length variants.
2. Assess our data in association with all of the recorded data using meta-analysis.

3. Patients and materials

3.1 Patients and Controls

All patients in this study were Norwegian, with a mean age of 68.2 years and similar number of male and female. The controls were the patient's spouse or one of their relatives that were neurologically healthy; controls had a mean age of 67.4 years. The study group consisted of 191 patients with PD (mean age of 68.2 years) and 191 age-matched, healthy controls from the Norwegian ParkVest study, a prospective, population based study of incident PD in Western and Southern Norway. A detailed description of the ParkVest study design has been published (27).

My clinical colleagues have ensured that the baseline examination including the Unified Parkinson Disease Rating Scale part II and III (subunits for Activities of Daily Living and motor symptoms) and the Hoehn and Yahr scale (HY) (69) was performed at the time of initial diagnosis with follow-up visits every half year. The motor subtype was defined as Postural Instability and Gait Difficulty (PIGD), indeterminate, or Tremor Dominant (TD) depending on the dominating motor symptom (70).Only individuals fulfilling clinical diagnostic criteria for PD (71) after at least 3 years follow-up were included (n=195) and of these sequencing results were obtained from 191. Controls were recruited from the patient's spouses, friends and others. We excluded 1st and 2nd degree relatives of patients (n=10), one control person who developed clinical PD within the first three years of follow-up as well as those without valid sequencing results (n=2). After matching for age, the control group consisted of 191 individuals.

3.2 Material

Commercial Kits

Article	**Catalogue No.**	**Producer**
AmpliTaq Gold® DNA Polymerase	4338856	Life Technologies, Invitrogen
GeneAmp® 10X PCR Gold Buffer	4338856	Life Technologies, Invitrogen
25 mM MgCl2	4338856	Life Technologies,

		Invitrogen
2.5 mM dNTP Mix	R725-01	Life Technologies, Invitrogen
Betaine	070M6047	
BigDye® Terminator v1.1 Cycle Sequencing Kit	4336772	Life Technologies, Invitrogen
GeneJET™ PCR Purification Kit	#K0702	Fermentas GmbH
QIAamp DNA Mini Kit	51306	QIAGEN
Sequence Buffer	4336697	Life Technologies, Invitrogen

Primers for PCR

Target gene	**Primer sequence**
*POLG*1	Fw 5' – GGT GGA AGG CAG GCA TGG TCA AA – 3'
	Rev 5' – GCT GGT CCA GGT TGT CCC CGT AGA G – 3'

General chemicals and solutions

Chemical/Solution	**Catalogue No.**	**Supplier**
Ethidium Bromide (EtBr)	E1510	Sigma-Aldrich, USA
Seakem LE Agarose		
6X Orange DNA Loading Dye	R0631	Thermo Fisher Scientific Inc
GeneRuler™ 50 bp DNA Ladder, ready-to-use	SM1133	Thermo Fisher Scientific Inc
10X TBE solution	Am9863	Life Technologies, Ambion

Hardware

Name	**Catalogue No./Model**	**Supplier**
NanoPhotometerTM P-Class	P 300	Implen GmbH, Germany

Thermal Cycler	Veriti	Applied Biosystems

Software

Program	**Version**	**Supplier**
Microsoft Office	2007	Microsoft Corporation , Redmond, USA
Graphpad	QuickCalc free web calculator	GraphPad Software, Inc, CA 92037 USA
SPSS	19	SPSS Inc., USA.
Geneious	5	Biomatters Ltd, New Zealand

3.3 Methods

3.3.1 DNA extraction, amplification and sequencing

DNA was extracted from frozen (-80C) whole blood using a commercial kit (QIAamp DNA Mini Kit, QIAGEN) according to the manufacturer's protocol. DNA concentration and quality was determined using a Nano Photometer P300 spectrophotometer (Implen GmbH, Germany).

The polymerase chain reaction (PCR) is a robust technique to amplify DNA *in vitro* (72).PCR comprises three stages: melting DNA to separate the double strands (above 90°C), annealing of the primer (temperature depends on the primer sequence usually ~55°-68°C) and DNA synthesis: extension (depends on DNA polymerase enzyme 72°C). The PCR reactions were carried out using a Veriti thermal cycler (Applied Biosystems, USA). Total volume of each reaction was 25 microliters (µl) and each reaction contained 1.5 mM $MgCl_2$ (Invitrogen # 4338856), 0.8 mM dNTP (Invitrogen # R725-01), 1.50 M betain (Invitrogen # 070M6047), 0.6 unit (U) AmpliTaq polymerase (Invitrogen # 4338856), 1x PCR-Gold buffer (Invitrogen # 4338856), 0.6 micro molar (µM) of each primers (forward and reverse) and 100 nano gram (ng) genomic DNA as template. The combination of Milli-Q water and template varied according to concentration of the DNA for each sample, but each reaction were carried out in a total volume of 25 microliters (µl).

Total PCR cycles were 30 which each cycle consists of the DNA denaturation for 30 sec at 94 °C, the primer annealing for 30 seconds at 58 °C and the elongation for 30 seconds at 72 °C.
PCR products were run on 1% agarose gel containing ethidium bromide (0.5um/ml) at 40V for 120min. Amplicon purification was done with GeneJET™ PCR Purification Kit (#K0702, Fermentas GmbH) according to the manufacturer's protocol.

3.3.2 Sanger Sequencing (Chain-termination methods)

Sequencing was carried out with BigDye® Terminator v1.1 Cycle Sequencing Kit (ABI) in a 10 µl reaction which contained 1 µl BigDye® Terminator v1.1 (ABI, cat# 4336772), 2 µl sequencing buffer (5X) (ABI, cat# 4336697), 0.5 µl primer (5 pmol), 2 µl betain (ABI, cat# 070M6047) and 2 µl DNA template (1-10 ng). Sequence of all samples was performed using forward primer that was used in PCR except the samples with low quality of chromatogram which reversed primer used, too. The products were analyzed on an ABI 3100 automated DNA sequencer.

Total PCR cycles for sequencing were 25 and each cycle consist of 10 sec DNA denaturation at 96 °C for , 5 seconds primer annealing at 50 °C and 4 minutes for elongation at 60 °C .

3.3.3 Analyze sequences (Bioinformatics)

Heterozygous *POLG1* sequence data were analyzed with the Geneious software v5.1 (http://www.geneious.com) and the NCBI sequence NM_002693.2 was used as a reference for *POLG1*.

3.3.4 Statistical analysis

To evaluate the differences in frequency of the genotypes between the patient and controls and find out any possible association between any specific allele and PD Fisher exact and chi-square test were applied. Fisher's exact test, always gives an exact P value and works fine with small sample sizes, so we used it for our result from Norwegian population to find out is there enough evidence to reject the null hypothesis which was "There is no association between length of CAG tract in *POLG*1 gene and Parkinson disease"; however the chi-square test gives the

appropriate result for large sample sizes. Chi-squared test was applied for meta-analysis to determine the difference of allele frequencies of *POLG*1 in combined PD and control group were significantly different and find the possible association between the frequencies of specific allele and PD. The p-value <0.05 was considered to indicate statistical significance. Odds ratio were estimated with 95 percent confidence intervals (CI 95%) in these studies.

With the use of simple 2 x 2 tables we can quantify the relationship between exposure and disease with odds ratio (OR). General structure of a 2×2 table:

General Structure of 2 × 2 table		*B* (Disease)		
		PD (+)	Control (-)	Totals
A (Exposure)	high exposure in PD group (+)	a	b	a+b
	high exposure in control group (-)	c	d	c+d
Total		a+c	b+d	N=a+b+c+d
$p_A = (a+b)/N$ $p_B = (a+c)/N$				

Odds ratio formula based on 2×2 table:

$$OR = \frac{a \times d}{b \times c}$$

4 Results

4.1 Higher frequency of common *POLG1* alleles in the control group compared to the PD group

Table 1 shows the *POLG1* allele frequency in patients and controls and table 2 shows the genotype frequencies. Figure 6 shows the allele distributions in control group and PD group

Allele	PD, n	PD %	Control, n	Control %	Total	Total %	P-Value
3R(CAG)8	4	1.05	1	0.26	5	0.65	0.373
3R(CAG)9	5	1.31	2	0.52	7	0.92	0.451
3R(CAG)10	297	77.75	322	84.29	619	81.02	0.027
3R(CAG)11	50	13.09	40	10.47	90	11.78	0.313
3R(CAG)12	23	6.02	10	2.62	33	4.32	0.031
3R(CAG)13	3	0.79	4	1.05	7	0.92	1.000
3R(CAG)14	0	0.00	1	0.26	1	0.13	1.000
3R(CAG)16	0	0.00	1	0.26	1	0.13	1.000
4R(CAG)9	0	0.00	1	0.26	1	0.13	1.000
Total	382		382		764		

Table 1. Allele frequencies of POLG1 in control group and patient group. Last column shows the P-value which calculate with Fisher exact test.

Genotype	PD (n)	PD %	Control (n)	Control %	Total	%
8/10	3	1.57	1	0.52	4	1.05
9/10	5	2.62	1	0.52	6	1.57
10/10	113	59.16	136	71.20	249	65.18
11/10	39	20.42	34	17.80	73	19.11
12/10	21	10.99	7	3.66	28	7.33
13/10	3	1.57	4	2.09	7	1.83
14/10	0	0.00	1	0.52	1	0.26
15/10	0	0.00	0	0.00	0	0
16/10	0	0.00	1	0.52	1	0.26
8/11	1	0.52	0	0.00	1	0.26
9/11	0	0.00	1	0.52	1	0.26
11/11	4	2.09	1	0.52	5	1.31
12/11	2	1.05	3	1.57	5	1.31
4R+9Q/10	0	0	1	0.52	1	0.26
Total	191		191		382	

Table 2. **Genotype frequencies for *POLG1***. 382 samples in two groups (control and PD) analyzed. 10/10 was the most frequent genotype in both groups.

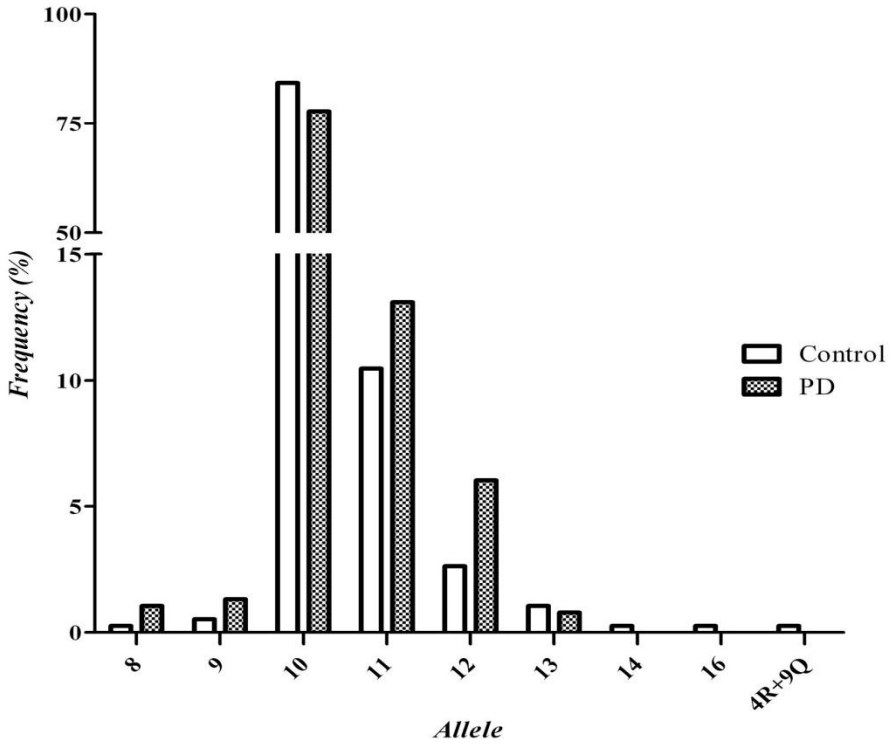


Figure 3. **Allele frequencies**. (A) distribution of POLG1 in control group and PD group in Norwegian population. In both group common allele was 10Q (PD; 297 Control; 322) and next common allele after 10Q was 11Qin both groups (PD; 50 Control; 40)

Irrespective of CAG repeat length, there was a higher proportion of heterozygous individuals in the PD group compared to the control group, however the control group had a higher number who were homozygous (see figure 7). In total (i.e. all 382 individuals), 66.5 % were homozygous (10/10 or 11/11) of which 65.2 % were 10/10 (59.2 % of the PD and 71.2 % of the control group) and 1.3 % were 11/11 (2.1 % of the PD group and 0.5 % of the control group). In contrast, 33.5 % were heterozygous (38.7 % PD group and 28.3 % control group) with 93.8 % of the heterozygous s having one 10-repeat allele and a second allele with a repeat length of 8-16 (55.5 % PD and 38.3 % Control group).

The percentage of individuals who were heterozygous for their CAG repeats length was significantly higher in the PD group compared to control group (P= 0.0392; OR 1.25, 95% CI 1.04 to 2.46). Three percent (3%) of the samples had no 10 CAG repeat allele (3.7% PD and 2.6% Controls) and the only 4R variants were in the control group (one allele 3R (CAG) 10 and another one 4R (CAG) 9).

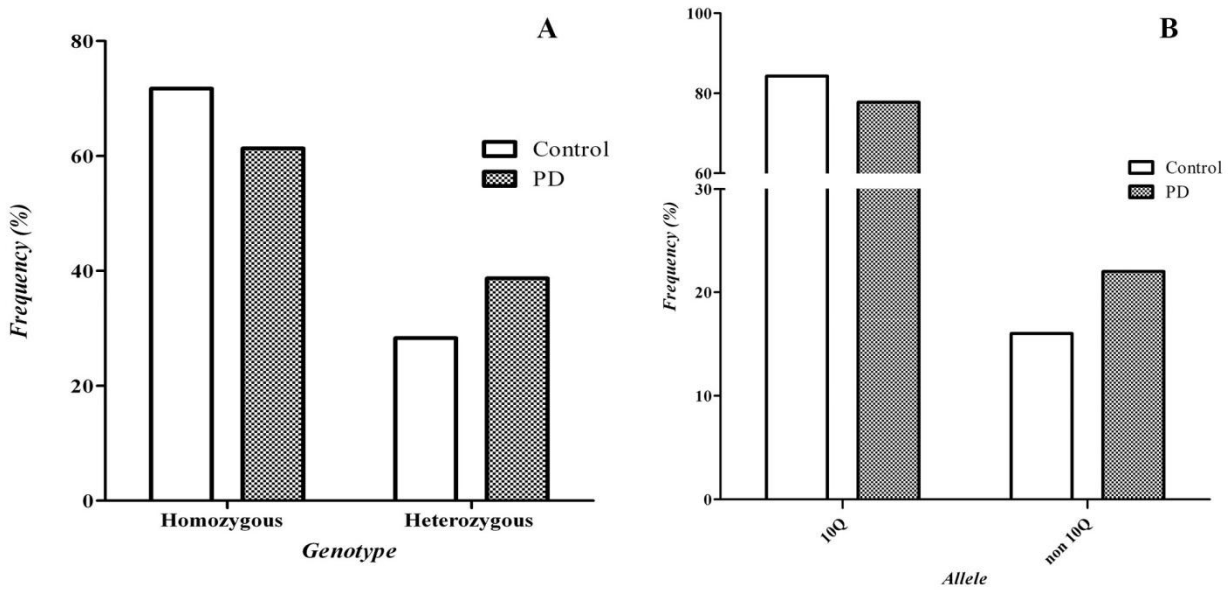


Figure 7. (A) Genotype frequencies in PD group and control group. Most common genotype in both group were 10/10. High level of heterozygosity in PD group compare to control group (B) Frequencies of common allele (10Q) and rare alleles (non 10Q) in control and PD group. High level of non 10Q alleles in PD group compare to control group.

Analysis of DNA sequences from both groups (patients and controls) revealed that the ten-repeat allele (10Q) is the most common allele in the Norwegian population; the frequency of this allele in the control group (84.3 %) was higher than PD group (77.7 %). There was also a significant difference between numbers of non 10Q alleles in the PD group compared to control (P= 0.027; OR 1.53, 95% CI 1.06 to 2.21). The most common allele after 10Q was 11Q and this was true for both groups: in total, 20.4 % of all samples had one 11Q allele (22.5 % PD group and 18.3 % Control group) and 1.3 % were homozygous for 11Q (2.1 % PD group and 0.5% Control group).

The frequency of both the 11Q and 12Q alleles was higher in PD group (11Q; 13.1 %, 12Q; 6.0 %) than the control group (11Q; 10.5 %, 12Q; 2.6 %); however, the difference was significant only for the 12Q alleles in PD group and control group (P= 0.031; OR 2.38, 95% CI 1.11 to 5.07). Two rare alleles, the 14Q and 16Q, were found in the control group and both individuals were heterozygotes with a 10Q on the other allele. There were no homozygotes for rare alleles (8, 9, 12, 13, 14 and 16Q) in our samples.

Based on the Hardy-Weinberg equation, it can be predicted that 10/10 alleles should occur in approximately 66 % of individuals in whole population (patient and control groups). In our

study, we found 65.2 % carried the 10/10 genotype, which approximates well to the prediction. Looking at other allele combinations, we found that 10/X (X is any allele except 10) was 33.2 % as opposed to 30.6 % predicted, and for X/X we found 3.1% versus 3.6 % predicted.

4.2 Meta-analysis: Non-10Q alleles correlate with PD

In this pooled study we used all available previous studies that had investigated the association between CAG trinucleotide repeat length in the *POLG*1 and Parkinson disease (16, 63, 65, 67, 68, 73). Figure 8 shows the distribution of alleles of the combined control and PD groups. All of these studies had three criteria: 1) Control subjects were neurologically healthy individuals; 2) The patients were diagnosed as idiopathic PD; 3) Number of CAG repeat in each allele among the cases and the controls reported. In total there were 1597 patients with a mean age of 69.2 and 1493 controls with a mean age of 67.4. The populations were from England, Sweden, Finland, United States and Norway. (Table 1. in supplementary data)

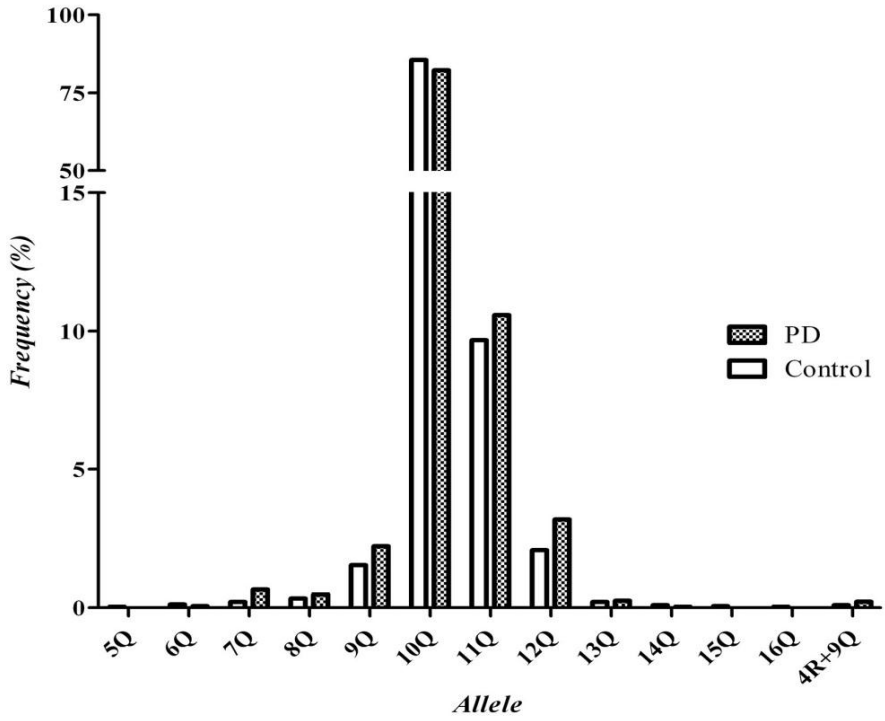


Figure 8. **Allele frequencies.** Allele frequencies of combined control and PD groups. The most common allele in control and PD group was 10Q (PD; 82.31 %, control; 85.54 %). The frequency of the 10 Q in PD group (82.31 %) was lower than control group (85.54 %)

Poly-Q	combined PD	combined PD %	combined Control	combined Control %	Total	Total %
5Q	0	0.00	1	0.03	1	0.0%
6Q	2	0.06	4	0.12	6	0.1%
7Q	21	0.66	7	0.21	28	0.4%
8Q	15	0.47	11	0.33	26	0.4%
9Q	71	2.22	52	1.54	123	1.9%
4R+9Q	7	0.22	3	0.09	10	0.2%
10Q	2629	82.31	2881	85.54	5510	84.0%
11Q	338	10.58	326	9.68	664	10.1%
12Q	102	3.19	70	2.08	172	2.6%
13Q	8	0.25	7	0.21	15	0.2%
14Q	1	0.03	3	0.09	4	0.1%
15Q	0	0.00	2	0.06	2	0.0%
16Q	0	0.00	1	0.03	1	0.0%
Total	3194		3368		6562	

Table 3. **Frequencies of *POLG1* alleles in meta-analysis**. *POLG1* allele frequencies in a mixed group of control and PD samples compromised of 7 studies.

Once again, the most common allele in both control and PD group was 10Q (PD; 82.31 %, control; 85.54 %). The frequency of the 10Q is lower in the PD group than control group, however, the frequency of non 10Q allele (7, 8, 9, 11, 12,13Q) was higher in the PD group than controls (see table 3). There was also a significant association between non 10Q alleles (≠10Q) and PD ($\chi2$=12.70, 1 df, P= 0.0004; OR 1.27, 95% CI 1.11 to 1.45). The second commonest allele in both groups was 11Q (PD; 10.58 %, control; 9.68 %). The frequency of the 11Q and 12Q allele in PD group (11Q; 10.58 %, 12Q; 3.19 %) was higher than control group (11Q; 9.68 %, 12Q; 2.08 %) and there was significant association between 12Q of CAG and PD ($\chi2$=7.98, 1 df, P= 0.0047; OR 1.55, 95% CI 1.14 to 2.11).

9Q that had high frequency in PD group (2.22 %) compared to control group (1.54 %) which the difference was significant according to Chi-square test ($\chi2$=4.109, 1 df, P= 0.0427; OR 1.21, 95% CI 1.01 to 2.08).

Odds ratio calculated for each study with 95% confidence intervals (95% CI) to estimate probability of increasing risk of PD in persons with non 10Q allele (figure 9). In conclusion overall odds ratio was OR=1.21 (95% CI) which was lower than our study for Norwegian population (OR=1.53, 95 CI).

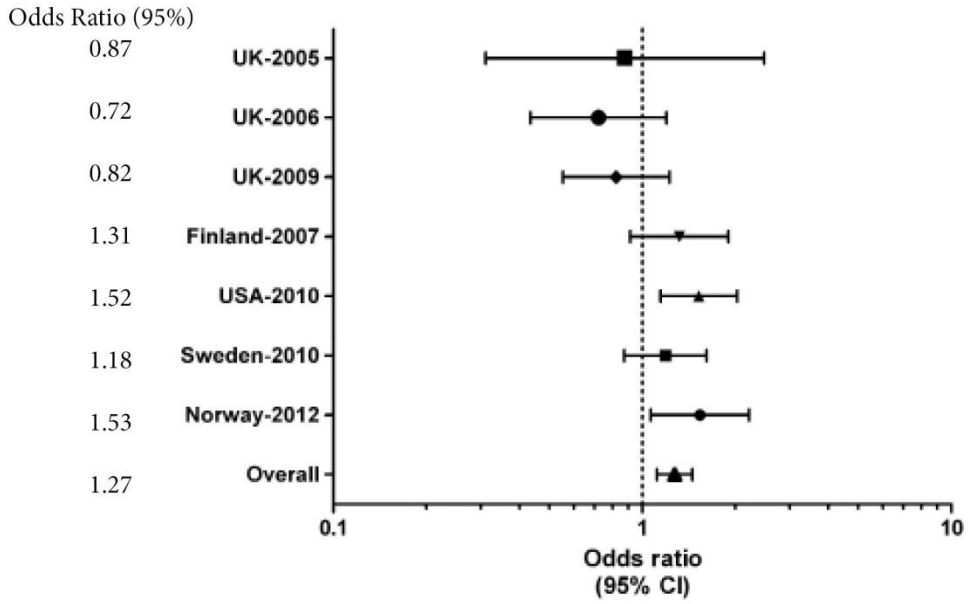


Figure 9 . Forest plot for the odds ratio (95% CI) for association between non 10 Q repeats and PD. The X axis show a Log 10 scale and each study has three values (the odds ratio, and the lower and upper confidence limits).

5 Discussion

5.1 The main experimental findings

The main findings of this study were:

- The most common allele in both the control and PD groups was the 10Q.
- 11Q was the second most frequent allele in both groups.
- A high frequency of heterozygosity (10/X) in the PD group compared to the controls.
- A high frequency of homozygosity (10/10 and 11/11) in the control group compared to patients with PD.
- A high frequency of non 10Q alleles in PD group, especially 12Q.
- A significant association between non 10Q (P=0.027) and Parkinson disease.
- A significant association between 12Q (P=0.031) and Parkinson disease.
- The most common alleles in the meta-analysis of controls and PD patients were 10Q followed by 11Q.
- Meta-analysis also showed a significant association between 9Q (P=0.0427), 12Q (P=0.0047) and all non 10Q (P= 0.0004) alleles and Parkinson disease.

In order to investigate the possible role of the CAG repeat length variation at *POLG1* gene in Parkinson disease we have studied repeat length variation at this locus in both patients with PD and age-matched controls. All individuals were Norwegians and there similar number of both genders. Also In addition we performed a meta-analysis of all previously published results that have looked at CAG length polymorphism in the *POLG*1 gene and PD (6 published articles until 2010). A total of 1597 PD cases and 1493 controls were found.

5.2 The biology of the CAG repeat in *POLG1*

Our results show that 10Q is the most common allele both in the Norwegian population (81 %) and the pooled analysis (84 %). Interestingly, this is species specific Gorilla is similar to human with 10Q being the commonest, but in chimpanzees it is 7Q and bonobos 5Q. Some species such as the mouse and drosophila lack a CAG tract in *POLG1* altogether (74). Several studies showed that 10Q allele is the most common allele in general population of humans (58, 59, 64, 65). The next most common allele in our and all previous studies was 11Q; together the 10Q and 11Q

alleles comprised over 90 % in our study (92.8 %) and the pooled analysis (95.22%). This appears to be an evolutionary conserved principle: primate species have one common allele for *POLG1* and the next common allele in each of these species has one more glutamine and these two alleles (most common allele and the second most common allele which has one more amino acid) comprised more than 90 % of allele frequencies (74).

The finding of a range of TNR in a specific gene in general population appears to be a normal phenomenon. The range can vary and can be large (11-34 TNR in the Huntington's disease gene or 6-39 in SCA, etc. (75)). The range in *POLG1* appears much narrower with 10 CAG repeats being by far the most prevalent allele in general population irrespective of whether they had PD or not. The range of CAG repeats in *POLG1* can vary between 5Q to 16Q, but given the predominance of the 10Q (or 11Q), it is possible that optimal Polγ functions depends a specific number of glutamine amino acid and that variations of this number can lead to variation in Polγ function. Impairment of Polγ function can potentially cause several problems for mitochondria including deletions of mtDNA and mtDNA depletion. These in turn can lead to respiratory chain deficiency. Indeed abnormal respiratory chain function has been demonstrated in PD patient's brain at autopsy (39, 43) as have mtDNA deletions.

The only non 10Q that we found in the homozygous state in our study was 11/11 (2.1 % PD group and 0.5 % Control group) and this also showed a higher frequency in PD group compared controls. The lack of homozygous non 10 or 11Q generally, and higher frequency of the only non 10Q homozygous in PD group, suggest that allele length might indeed be important for optimal Polγ performance.

5.3 Lack of the common 10Q allele in *POLG1* is associated with PD

In this study, we classified the 10Q allele as the common allele and all the other alleles were grouped as rare alleles. In some other studies, 10Q and 11Q alleles were grouped together for statistical analysis. We did not do this. **We found a significant association between non 10 CAG repeat lengths and Parkinson diseases**, confirming the previous studies in Finland, Sweden and the United States (16, 63, 65). Of the non-10/11Q variants, the strongest association was found with 12Q and this is also confirmed by our meta-analysis. Another allele with a significant association with PD in the pooled analysis was 9Q. However, we did not find any significant association between 9Q and PD in our study, probably because of the sample size.

The 9Q was, however, the most frequent allele after 12 Q in the PD group and this was true for the other studies, although not the Finnish, in which 9Q was more frequent than 12Q. The Finnish population is characterized by isolated communities and low genetic diversity population which perhaps explains the difference between Finland's study and other studies (76).

The absence of an association between non 10Q and PD in the three studies from the UK may have different explanations. A small sample size is one [7]. In this study, the PD group consisted of 22 patients and 31 controls (73). In another, 140 cases of PD and 90 controls were studied (67)). In the study by Tiangyou, the samples were not matched for age and sex. Another important confounding factor is ethnicity; this is potential problem in the Finnish study (16) due to the low genetic diversity. Since the association between non 10Q and PD has been confirmed by several different studies in Nordic countries, the question arises of whether this is specific to these populations. However, the finding of an association between non 10Q and PD in the study from USA that investigated a mixed Caucasian control group would tend to refute this (65). Lastly, it is important to remember that technical problems may also play a role. For example, the fragment analysis that was used in the study by Anvret et al., 2010 cannot recognize a 9Q+1R (rs28567406) allele or any other possible sequence variation.

The observed homozygosity values (10/10 and non 10/non10) are close to equilibrium values predicted by the Hardy Weinberg equation. However, there was a slight difference between the Hardy–Weinberg predication for heterozygosity (10/X) in our study (prediction 30.6 %, our study 33.2 %), although this deviation was not statistically significant. These result shows that locus is not an interesting locus for evolutionary evaluation, and that the general population is at equilibrium. (Allele frequencies of 10 and X do not change from one generation to another)

5.4 Possible somatic instability of CAG repeats in *POLG*1 gene

Existing studies of the CAG polymorphism in *POLG*1 gene have usually compared the frequency of these variants in blood samples from controls and PD patients. Only one study (Taanman et al. 2005) studied CAG polymorphism in 6 post-mortem brains. On the basis of other CAG repeat disorders, somatic instability of the *POLG1* CAG length polymorphism in different tissues should also be considered. Gametic and somatic mosaicism of CAG repeat in the *SCA1* gene was shown by isolated single cell study looking at sperm, blood and brain tissue

of spinocerebellar ataxia type 1 patients (77). These studies showed shorter CAG repeats in the cerebellum, which has a lower rate of cell division during neuronal network development, and longer CAG repeat in other parts of the brain suggesting that differences in early mitotic activity may result in tissue and cell specific segregation of different CAG repeat lengths (77). Hashida et al. 2001 showed heterogeneity of CAG repeat expansion in Purkinje and cerebellar granule cells in late onset dentatorubral-pallidoluysian atrophy (DRPLA) patients. Their study also showed higher CAG repeat heterogeneity in glial cells. Mitotic rate of cell division can be a possible explanation for such a difference (78). The neuronal population is controlled by apoptosis of damaged neurons during neuronal network development, which may explain the survival of neurons with lower expansion and the finding of neurons with lower expansion in homogenized tissue (78). Extension of CAG repeat length in *POLG1* during mitotic division, especially during embryonic development by whatever mechanism (49) could lead to somatic mosaicism. This would make it difficult to assess the real proportion of existing alleles in tissues such as brain using conventional methods. According to these data, whole blood may not be a good tissue to analysis CAG length and any possible association with disease. In addition, any increase in the level of somatic mosaicism of *POLG1* polymorphism heterogeneity during the human lifespan raises the possibility of finding the CAG repeat with longer length. At present, however, there is no evidence of either somatic differences of *POLG1* CAG length or changes with time.

5.5 Function of CAG repeat in *POLG*1 gene and Pol γ

In silico analysis of RNA-folding shows that there is a 4.8% increase in required folding energy for variants with more than 10/11Q. The difference between the minimum free energy structure of the 10/11Q and non 10/11Q sequences may give different secondary structures of mRNA and differing stability (63). Interestingly, Spelbrink et al. reported that *POLG1* expression is modestly up-regulated *in vitro* by removing the CAG repeat segment (modified *POLG1* by deletion 10 CAG repeats of *POLG1* gene in in cultured HEK293T cells). In addition, their study indicated that there is no difference in enzymatic activity of modified *POLG1*. However, this study did not elaborate how the CAG tract may modulate *POLG1* expression or enzymatic activity (18). This experiment indicated CAG repeats in *POLG1* gene can regulate the expression of Pol γ; however the functional effect of the CAG tract length polymorphism remains unclear.

Kozlowski et al., 2010 showed a significant association between TNRs and the regulation of gene expression, transcription, DNA binding and transcription regulator activity in a functional association analysis (47). These data offer a potential explanation for the result of Spelbrink's study which showed that loss of the CAG tract was associated with alteration of *POLG1* expression *in vitro.* These all data together suggest that the CAG TNR in *POLG1* may have a regulatory function for expression of *POLG1*.

Butland et al., 2007 also showed a positive correlation between the length of the CAG TNR and the level of polymorphism. They indicated that genes with uninterrupted CAG repeats of 10 or longer can potentially expand and become be a candidate gene for polyglutamine expansion disorders (79). However the study by Alba et al, 2001 suggested that CAG tracts interrupted by a CAA are more stable compared to uninterrupted tracts of equal length. The *POLG1* CAG repeat is normally followed by CAA and perhaps this is why the 10Q allele is so prevalent in the normal population (80).

Another important factor concerning the *POLG1* CAG repeats, is the location of this TNR. The polyglutamine tract comes just after the mitochondrial import signal (81). Polymerase γ is translocated into the matrix of mitochondria by passing through the TOM40 complex (the translocase complex of the outer mitochondrial membrane) and TIM23 complex (the translocase complex of the inner mitochondrial membrane) (82). The importance of the CAG repeat region in *POLG1* and its role in protein binding or protein interaction are unknown, but are potential sites of dysfunction. For example, expansion of this region might disrupt interaction between chaperon and import signal. Nasir .et al. 1996 showed that expansion of the polyglutamine tract in Huntingtin can lead to a conformational change causing altered protein mobility. In tissues such as brain and nerve cells which have a high energy demand any alteration in protein conformation which affects the Polγ translocation could potentially affect mitochondria biogenesis leading to decreased numbers of mitochondria or mtDNA depletion.

5.6 Role of *POLG1* polymorphism in mtDNA deletion and causation of PD

The prevalence of PD increases with age over 50 years and it is significantly higher between 65 and 90 years (83). There are several studies which have shown that mtDNA deletions increase during human lifespan in both muscle and neurons. The mechanisms that keep the balance

between the number of normal and deleted mtDNAs during a lifetime have yet to be identified (84), but this balance is more pronounced in tissues such brain and muscle probably due to their lack of regenerative capacity. There is evidence that mtDNA deletions accumulate during the human lifespan (84) and that PD patients accumulate deletions of mtDNA in the substantia nigra (85). *In silico* and *in vitro* studies suggest that altering CAG length in *POLG1* can influence the mRNA stability and expression of *POLG1*. This could in turn affect Pol γ function. Thus it is appropriate to speculate that a dysfunctional Polγ can potentially affect aged nerve cells more than other tissues in this age range and lead them to lose more mtDNA. This can cause respiratory chain deficiency and lead the cell to death.

5.7 *POLG* polymorphism can be a genetic marker

Linkage and association should not be confused with one another. Two genetic loci are linked if they are inherited from parents to offspring more often by the chance and these two genetic loci are in linkage disequilibrium (LD) if they are found together on the same haplotype in population is greater than expected by chance (86). In fact, Linkage disequilibrium shows the nonrandom association between alleles at different loci (87). LD gives us information about past events and causes of gene-frequency in population such as natural selection during the time, gene conversion and mutation (88). LD study showed the relation between CAG and GGC microsatellites in Androgen Receptor gene (89).

Genome Wide Association Studies (GWAS) have not shown an association between chromosome 15 (where *POLG1* is located) and Parkinson's disease (90). This does not exclude however a possible role for *POLG* sequence variation in PD as GWAS studies are generally less sensitive for rare variants. (91). Also, POLG1 haplotype analysis indicated six POLG haplotypes which there were not significant association between them and PD, however 9Q-CCAG haplotype had higher frequency among the PD group to compare with spouse controls (4.9% vs. 1.3%). The small size of the samples can be the reason for non-significant association, but still their finding support our result from meta-analysis (31).

Alternatively, it is possible that allelic variation of *POLG1* itself does not influence PD, but acts as a marker for other, nearby localized genetic variants which increase susceptibility for PD and

are present in linkage disequilibrium with one or more of non-10/11Q alleles in certain populations.

6. Conclusion

Our study showed that non10 repeats in the *POLG1* gene were more frequent in PD patients than in healthy population. This was especially true for the 12Q repeat. We also found the converse, that 10Q was significantly associated with control individuals. The 9Q allele, another polymorphism in the *POLG* gene, was significantly associated with PD in the pooled analysis but this was not confirmed in our study. Our data confirms previous studies and suggests that there is an association between CAG length polymorphism on the POLG gene and PD.

The role of the CAG repeats in *POLG*1 gene is still unclear, but we speculate on the basis of *in silico* and *in vitro* studies, that altering CAG length affects Pol γ activity, which can lead to failure of the enzyme to properly replicate mtDNA and can therefore generate mtDNA deletion or depletion. If however, non 10Q repeats, particularly 12Q repeats are simply a marker for Parkinson disease the mechanism(s) underlying this association remain obscure.

7. Future perspectives

In the future, further studies on polymorphic CAG repeat in *POLG1* in other geographical regions and with powerful molecular methods such as X-ray crystallography, single cell studies, in vitro studies and Genome Wide Association study (GWAS) may help to find the possible role of *POLG1* polymorphism in the CAG region in mitochondrial dysfunction and Parkinson disease.

There are several parameters involved in mRNA stability and functions prediction by bioinformatics software which makes a complicated process with complex results that explanation of these result needs expertise. One of the criteria of mRNA prediction is Folding Free Energy (ΔG) to find out which conformation of mRNA is stable or has higher chance to be in the environment. Mfold web server (92) is well-known program to estimate the folding free energy (ΔG) for mRNA. Anvret et al., 2010 showed the Folding Free Energy difference between the 10Q and non 10Q, so this time we used this program to find out the difference between folding free energy in 10Q allele and 12Q allele ((ΔG10- ΔG12)= 0.12%) (Figure 1, supplementary data) which showed slightly difference, although this program has good algorithm to predict second structure for small RNA such as tRNAs and rRNAs, but the performance of this program decrease with size of the RNAs (93), so the real difference can be higher than this. Second structure prediction for *POLG* protein was the other important feature that we looked at to find any possible difference between *POLG* polymorphism (performed by CLC Main Workbench 6.6.2, CLC Bio). There was no difference between 10Q and 11Q, but one β sheet elicited from Pol γ with 12Q. These finding can be an explanation for finding high frequencies of 12Q allele in patients group (Table 2, supplementary data).

Using single cell studies to investigate somatic instability of CAG repeat in *POLG*1 gene in different cell types and different tissue with different age onset which can be useful to understand the possible expansion of *POLG* gene during lifespan and shows possible link between *POLG* polymorphism and mitochondria dysfunction in substantia nigra.

Modified *POLG* gene and cloning methods such as study of Spelbrink et al 2000 for assessing or predicting the effects of CAG length in *POLG* gene expression and Pol γ function, could be

investigated in vitro to able us find out the other property of the CAG repeat polymorphism function.

Further investigations including association studies in other and larger populations can help us to find out the possible relation between CAG repeats in *POLG1* and other mutation in *POLG* gene or possible linkage disequilibrium with the other gene or mutations on the other neighbor genes.

8. References

1. Cooper GM, Hausman RE. The Cell: A Molecular Approach: Sinauer Associates; 2007. 1-820 p.
2. Lang BF, Burger G, OKelly CJ, Cedergren R, Golding GB, Lemieux C, et al. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature. 1997;387(6632):493-7.
3. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806):457-65.
4. Chen H, Chan DC. Mitochondrial Dynamics in Mammals. In: Gerald PS, editor. Current Topics in Developmental Biology: Academic Press; 2004. p. 119-44.
5. Perkins G, Renken C, Martone ME, Young SJ, Ellisman M, Frey T. Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. J Struct Biol. 1997;119(3):260-72.
6. Scheffler IE. Biogenesis of Mitochondria. Mitochondria: John Wiley & Sons, Inc.; 2007. p. 60-167.
7. Wikström MrKF. Mitochondrial energy transduction and oxidative phosphorylation. In: Holt IJ, editor. Genetics of Mitochondrial Diseases: Oxford University Press; 2003.
8. Hood DA. Plasticity in skeletal, cardiac, and smooth muscle - Invited review: Contractile activity-induced mitochondrial biogenesis in skeletal muscle. J Appl Physiol. 2001;90(3):1137-57.
9. Scheffler IE. Mitochondria2007. 484- p.
10. Brown TA, Cecconi C, Tkachuk AN, Bustamante C, Clayton DA. Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. Genes Dev. 2005;19(20):2466-76.
11. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806):457-65.

12. Chial HC, J. mtDNA and Mitochondrial Diseases. In: generation Trotmgie, editor.: Nature Education 2008.

13. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal Inheritance of Human Mitochondrial-DNA. P Natl Acad Sci-Biol. 1980;77(11):6715-9.

14. Fan L, Sanschagrin PC, Kaguni LS, Kuhn LA. The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: implications for a dual role as a primer recognition factor and processivity clamp. Proc Natl Acad Sci U S A. 1999;96(17):9527-32.

15. Bailey CM, Anderson KS. A mechanistic view of human mitochondrial DNA polymerase gamma: Providing insight into drug toxicity and mitochondrial disease. Bba-Proteins Proteom. 2010;1804(5):1213-22.

16. Luoma PT, Eerola J, Ahola S, Hakonen AH, Hellstrom O, Kivisto KT, et al. Mitochondrial DNA polymerase gamma variants in idiopathic sporadic Parkinson disease

10.1212/01.wnl.0000276955.23735.eb. Neurology. 2007;69(11):1152-9.

17. Copeland WC. Inherited mitochondrial diseases of DNA replication. Annu Rev Med. 2008;59:131-46.

18. Spelbrink JN, Toivonen JM, Hakkaart GAJ, Kurkela JM, Cooper HM, Lehtinen SK, et al. In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. J Biol Chem. 2000;275(32):24818-28.

19. Stumpf JD, Copeland WC. Mitochondrial DNA replication and disease: insights from DNA polymerase gamma mutations. Cell Mol Life Sci. 2011;68(2):219-33.

20. Larsson NG, Clayton DA. Molecular genetic aspects of human mitochondrial disorders. Annu Rev Genet. 1995;29:151-78.

21. Naess K, Barbaro M, Bruhn H, Wibom R, Nennesmo I, Döbeln U, et al. Complete Deletion of a <i>POLG1 Allele in a Patient with Alpers Syndrome

JIMD Reports - Case and Research Reports, 2012/1. In: Ssiem, editor.: Springer Berlin Heidelberg; 2012. p. 67-73.

22. Graziewicz MA, Longley MJ, Bienstock RJ, Zeviani M, Copeland WC. Structure-function defects of human mitochondrial DNA polymerase in autosomal dominant progressive external ophthalmoplegia. Nat Struct Mol Biol. 2004;11(8):770-6.

23. Van Goethem G, Dermaut B, Lofgren A, Martin JJ, Van Broeckhoven C. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nat Genet. 2001;28(3):211-2.

24. Naviaux RK, Nguyen KV. POLG mutations associated with Alpers' syndrome and mitochondrial DNA depletion. Ann Neurol. 2004;55(5):706-12.

25. Tzoulis C, Engelsen BA, Telstad W, Aasly J, Zeviani M, Winterthun S, et al. The spectrum of clinical disease caused by the A467T and W748S POLG mutations: a study of 26 cases. Brain. 2006;129(7):1685-92.

26. Parkinson J. An essay on the shaking palsy (Reprinted). J Neuropsych Clin N. 2002;14(2):223-36.

27. Alves G, Muller B, Herlofson K, HogenEsch I, Telstad W, Aarsland D, et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. J Neurol Neurosurg Psychiatry. 2009;80(8):851-7. Epub 2009/02/28.

28. von Campenhausen S, Bornschein B, Wick R, Botzel K, Sampaio C, Poewe W, et al. Prevalence and incidence of Parkinson's disease in Europe. Eur Neuropsychopharmacol. 2005;15(4):473-90.

29. Galindo MF, Jordan J, Gonzalez-Garcia C, Cena V. Chromaffin cell death induced by 6-hydroxydopamine is independent of mitochondrial swelling and caspase activation. J Neurochem. 2003;84(5):1066-73.

30. Blum D, Torch S, Lambeng N, Nissou M-F, Benabid A-L, Sadoul R, et al. Molecular pathways involved in the neurotoxicity of 6-OHDA, dopamine and MPTP: contribution to the apoptotic theory in Parkinson's disease. Prog Neurobiol. 2001;65(2):135-72.

31. Eerola-Rautio J. SPECT-aided diagnostics and genetic risk factors in Parkinson's disease: University of Helsinki; 2008.

32. Jr A, van Netten H, Schulzer M, Mak E, Mckenzie J, Strongosky A, et al. PET in LRRK2 mutations: comparison to sporadic Parkinson's disease and evidence for presymptomatic compensation. Brain. 2005;128:2777-85.

33. Samii A, Markopoulou K, Wszolek ZK, Sossi V, Dobko T, Mak E, et al. PET studies of parkinsonism associated with mutation in the alpha-synuclein gene. Neurology. 1999;53(9):2097-102.

34. Lim KL, Dawson VL, Dawson TM. The Cast of Molecular Characters in Parkinson's Disease. Annals of the New York Academy of Sciences. 2003;991(1):80-92.

35. Almeida MR. Glucocerebrosidase involvement in Parkinson Disease and other Synucleinopathies. Frontiers in Neurology. 2012;3.

36. Farrer M, Maraganore DM, Lockhart P, Singleton A, Lesnick TG, de Andrade M, et al. α-synuclein gene haplotypes are associated with Parkinson's disease. Hum Mol Genet. 2001;10(17):1847-51.

37. Schapira AHV. Mitochondria in the aetiology and pathogenesis of Parkinson's disease. Lancet Neurol. 2008;7(1):97-109.

38. Pathak RU, Davey GP. Complex I and energy thresholds in the brain. Bba-Bioenergetics. 2008;1777(7-8):777-82.

39. Davey GP, Peuchen S, Clark JB. Energy thresholds in brain mitochondria - Potential involvement in neurodegeneration. J Biol Chem. 1998;273(21):12753-7.

40. Janetzky B, Hauck S, Youdim MBH, Riederer P, Jellinger K, Pantucek F, et al. Unaltered Aconitase Activity, but Decreased Complex-I Activity in Substantia-Nigra Pars-Compacta of Patients with Parkinsons-Disease. Neurosci Lett. 1994;169(1-2):126-8.

41. Schapira AHV, Cooper JM, Dexter D, Jenner P, Clark JB, Marsden CD. Mitochondrial Complex I Deficiency in Parkinsons-Disease. Lancet. 1989;1(8649):1269-.

42. Bindoff LA, Birchmachin M, Cartlidge NEF, Parker WD, Turnbull DM. Mitochondrial-Function in Parkinsons-Disease. Lancet. 1989;2(8653):49-.

43. Schapira AHV, Cooper JM, Dexter D, Clark JB, Jenner P, Marsden CD. Mitochondrial Complex I Deficiency in Parkinsons-Disease. J Neurochem. 1990;54(3):823-7.

44. Esteves AR, Ardu, #237, no DM, Silva DFF, Oliveira CR, et al. Mitochondrial Dysfunction: The Road to Alpha-Synuclein Oligomerization in PD. Parkinson's Disease. 2011;2011.

45. Orsucci D, Ienco EC, Mancuso M, Siciliano G. POLG1-Related and other "Mitochondrial Parkinsonisms": an Overview. J Mol Neurosci. 2011;44(1):17-24.

46. Ellegren H. Microsatellites: Simple sequences with complex evolution. Nat Rev Genet. 2004;5(6):435-45.

47. Kozlowski P, de Mezer M, Krzyzosiak WJ. Trinucleotide repeats in human genome and exome. Nucleic Acids Res. 2010;38(12):4027-39.

48. Fischer KM. Parkinson's disease, amyotrophic lateral sclerosis and spinal muscular atrophy are caused by an unstable (CAG)n trinucleotide repeat microsatellite. Med Hypotheses. 1997;49(4):337-45. Epub 1997/11/14.

49. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. Annu Rev Neurosci. 2007;30:575-621.

50. Mirkin SM. Expandable DNA repeats and human disease. Nature. 2007;447(7147):932-40.

51. Nasir J, Goldberg YP, Hayden MR. Huntington disease: New insights into the relationship between CAG expansion and disease. Human Molecular Genetics. 1996;5:1431-5.

52. NCBI Reference Sequence. NCBI: NCBI Reference Sequence; 2012; Available from: http://www.ncbi.nlm.nih.gov/nuccore/NG_008218.1.

53. Ropp PA, Copeland WC. Cloning and characterization of the human mitochondrial DNA polymerase, DNA polymerase gamma. Genomics. 1996;36(3):449-58.

54. Azrak S, Ayyasamy V, Zirpoli G, Ambrosone C, Bandera EV, Bovbjerg DH, et al. CAG Repeat Variants in the POLG1 Gene Encoding mtDNA Polymerase-Gamma and Risk of Breast Cancer in African-American Women. Plos One. 2012;7(1).

55. Jensen MB, Leffers H, Petersen JH, Daugaard G, Skakkebaek NE, Meyts ERD. Association of the polymorphism of the CAG repeat in the mitochondrial DNA polymerase gamma gene (POLG) with testicular germ-cell cancer. Ann Oncol. 2008;19(11):1910-4.

56. Nowak R, Zub R, Skoneczna I, Sikora K, Ligaj M. CAG repeat polymorphism in the DNA polymerase gamma gene in a Polish population: an association with testicular cancer risk. Ann Oncol. 2005;16(7):1211-2.

57. Jensen M, Leffers H, Petersen JH, Andersen AN, Jorgensen N, Carlsen E, et al. Frequent polymorphism of the mitochondrial DNA polymerase gamma gene (POLG) in patients with normal spermiograms and unexplained subfertility. Hum Reprod. 2004;19(1):65-70.

58. Aknin-Seifer IE, Touraine RL, Lejeune H, Jimenez C, Chouteau J, Siffroi JP, et al. Is the CAG repeat of mitochondrial DNA polymerase gamma (POLG) associated with male infertility? A multi-centre French study. Hum Reprod. 2005;20(3):736-40.

59. Krausz C, Guarducci E, Becherini L, degl'Innocenti S, Gerace L, Balercia G, et al. The clinical significance of the POLG gene polymorphism in male infertility. J Clin Endocr Metab. 2004;89(9):4292-7.

60. Brusco A, Michielotto C, Gatta V, Foresta C, Matullo G, Zeviani M, et al. The polymorphic polyglutamine repeat in the mitochondrial DNA polyrnerase gamma gene is not associated with oligozoospermia. J Endocrinol Invest. 2006;29(1):1-4.

61. Harris TP, Gomas KP, Weir F, Holyoake AJ, McHugh P, Wu M, et al. Molecular analysis of polymerase gamma gene and mitochondrial polymorphism in fertile and subfertile men. Int J Androl. 2006;29(3):421-33.

62. Liu SY, Zhang CJ, Peng HY, Yao YF, Shi L, Chen JB, et al. CAG-repeat variant in the polymerase gamma gene and male infertility in the Chinese population: a meta-analysis. Asian J Androl. 2011;13(2):298-304.

63. Anvret A, Westerlund M, Sydow O, Willows T, Lind C, Galter D, et al. Variations of the CAG trinucleotide repeat in DNA polymerase gamma (POLG1) is associated with Parkinson's disease in Sweden. Neurosci Lett. 2010;485(2):117-20. Epub 2010/09/10.

64. Luoma PT, Eerola J, Ahola S, Hakonen AH, Hellstrom O, Kivisto KT, et al. Mitochondrial DNA polymerase gamma variants in idiopathic sporadic Parkinson disease. Neurology. 2007;69(11):1152-9.

65. Eerola J, Luoma PT, Peuralinna T, Scholz S, Paisan-Ruiz C, Suomalainen A, et al. POLG1 polyglutamine tract variants associated with Parkinson's disease. Neurosci Lett. 2010;477(1):1-5. Epub 2010/04/20.

66. Taanman JW, Schapira AH. Analysis of the trinucleotide CAG repeat from the DNA polymerase gamma gene (POLG) in patients with Parkinson's disease. Neurosci Lett. 2005;376(1):56-9. Epub 2005/02/08.

67. Tiangyou W, Hudson G, Ghezzi D, Ferrari G, Zeviani M, Burn DJ, et al. POLG1 in idiopathic Parkinson disease. Neurology. 2006;67(9):1698-700. Epub 2006/09/01.

68. Hudson G, Tiangyou W, Stutt A, Eccles M, Robinson L, Burn DJ, et al. No Association Between Common POLG1 Variants and Sporadic Idiopathic Parkinson's Disease. Movement Disord. 2009;24(7):1092-4.

69. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology. 1967;17(5):427-42. Epub 1967/05/01.

70. Jankovic J, McDermott M, Carter J, Gauthier S, Goetz C, Golbe L, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. Neurology. 1990;40(10):1529-34. Epub 1990/10/01.

71. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol. 1999;56(1):33-9. Epub 1999/01/29.

72. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific Enzymatic Amplification of DNA Invitro - the Polymerase Chain-Reaction. Cold Spring Harb Sym. 1986;51:263-73.

73. Taanman JW, Schapira AHV. Analysis of the trinucleotide CAG repeat from the DNA polymerase gamma gene (POLG) in patients with Parkinson's disease. Neurosci Lett. 2005;376(1):56-9.

74. Rovio A. DNA Polymerase Gamma Mutations in Male Infertility and Ageing [ACADEMIC DISSERTATION]. Tampere University Press: University Of Tampere; 2006.

75. Gatchel JR, Zoghbi HY. Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet. 2005;6(10):743-55.

76. Delachapelle A. Disease Gene-Mapping in Isolated Human-Populations - the Example of Finland. Journal of medical genetics. 1993;30(10):857-65.

77. Chong SS, Mccall AE, Cota J, Subramony SH, Orr HT, Hughes MR, et al. Gametic and Somatic Tissue-Specific Heterogeneity of the Expanded Sca1 Cag Repeat in Spinocerebellar Ataxia Type-1. Nat Genet. 1995;10(3):344-50.

78. Hashida H, Goto J, Suzuki T, Jeong SY, Masuda N, Ooie T, et al. Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA). J Neurol Sci. 2001;190(1-2):87-93.

79. Butland S, Devon R, Huang Y, Mead C-L, Meynert A, Neal S, et al. CAG-encoded polyglutamine length polymorphism in the human genome. Bmc Genomics. 2007;8(1):126.

80. Alba MM, Santibanez-Koref MF, Hancock JM. The comparative genomics of polyglutamine repeats: extreme differences in the codon organization of repeat-encoding regions between mammals and Drosophila. J Mol Evol. 2001;52(3):249-59.

81. Saneto RP, Naviaux RK. Polymerase Gamma Disease through the Ages. Dev Disabil Res Rev. 2010;16(2):163-74.

82. Endo T, Yamano K. Multiple pathways for mitochondrial protein traffic. Biol Chem. 2009;390(8):723-30.

83. VanItallie TB. Parkinson disease: primacy of age as a risk factor for mitochondrial dysfunction. Metabolism. 2008;57(10):S50-S5.

84. Simonetti S, Chen X, Dimauro S, Schon EA. Accumulation of Deletions in Human Mitochondrial-DNA during Normal Aging - Analysis by Quantitative Pcr. Biochim Biophys Acta. 1992;1180(2):113-22.

85. Bender A, Krishnan KJ, Morris CM, Taylor GA, Reeve AK, Perry RH, et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. Nat Genet. 2006;38(5):515-7.

86. Teare MD, Barrett JH. Genetic epidemiology 2 - Genetic linkage studies. Lancet. 2005;366(9490):1036-44.

87. Weir BS. Linkage disequilibrium and association mapping. Annu Rev Genom Hum G. 2008;9:129-42.

88. Slatkin M. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-85.

89. Irvine RA, Yu MC, Ross RK, Coetzee GA. The Cag and Ggc Microsatellites of the Androgen Receptor Gene Are in Linkage Disequilibrium in Men with Prostate-Cancer. Cancer Res. 1995;55(9):1937-40.

90. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009;41(12):1308-U68.

91. Pearson TA, Manolio TA. How to interpret a genome-wide association study. Jama-J Am Med Assoc. 2008;299(11):1335-44.

92. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406-15.

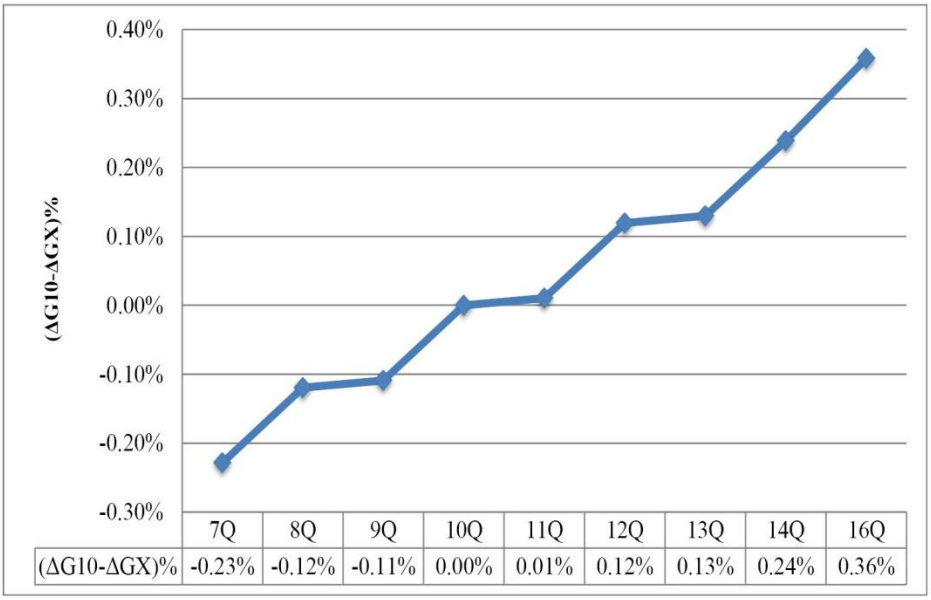
93. Zuker M, Jacobson AB. Well-Determined Regions in Rna Secondary Structure Prediction - Analysis of Small-Subunit Ribosomal-Rna. Nucleic Acids Res. 1995;23(14):2791-8.

9. Appendix

9.1Supplementary data

Table 1. Criteria of each study in Meta-analysis

	Ref	cohort	association	Case	Case's Gender	Case's mean age	Ctrl	Control's Gender	Ctrl's mean age	Origin
England	(68)	-	No	220	matched	71.41	157	matched	69.69	British (Caucasian)
England	(67)	UK cohort	No	140	61.4% male	70.97	90	36.7% male	69.19	British (Caucasian)
		Italian cohort	No	279	60.5% male	47	285	37.8% male	62	British (Caucasian)
England	(66)	-	No	22	matched	74	31	matched	71.6	British (Caucasian)
USA	(65)	Mixed Caucasian Ctrl	Yes	-	n.a.	n.a.	605	n.a.	n.a.	mixed Caucasian
		Main group	Yes	641	56% male	65.3	292	49% male	68.2	North American Caucasian
Finland	(64)	Main group	yes	140	83 Men	67.2	127	47 men	65.8	Finnish (Caucasian)
		unselected Ctrl	yes	-	n.a.	n.a.	516	n.a.	>85	Finnish (Caucasian)
		OND Ctrl	No	-	n.a.	n.a.	120	n.a.	n.a.	Finnish (Caucasian)
Sweden	(63)	-	Yes	243	n.a.	67.3	279	n.a.	59.9	Swedish (Caucasian)
Norway	2012	-	Yes	191	matched	68.2	191	matched	67.5	Caucasian



	7Q	8Q	9Q	10Q	11Q	12Q	13Q	14Q	16Q
(ΔG10-ΔGX)%	-0.23%	-0.12%	-0.11%	0.00%	0.01%	0.12%	0.13%	0.24%	0.36%

Figure 1. **Folding free energy predicted for second structure of POLG mRNA.** Illustrate Δ (ΔG) of POLG mRNA polymorphism base on different length of CAG in POLG gene. (Analysis by mFOLD program)

Allele	Alpha helices	Beta strands
POLG-6Q	50	23
POLG-7Q	50	23
POLG-8Q	50	23
POLG-9Q	51	22
POLG-10Q	51	22
POLG-11Q	51	22
POLG-12Q	51	21
POLG-13Q	51	21
POLG-14Q	51	21
POLG-16Q	51	21

Table 2. **Second structure prediction of POLG protein.** Determine number of Alpha helices and Beta strands in second structure of different POLG alleles. (Analysis performed by CLC Main Workbench 6.6.2, CLC Bio)