UNIVERSIDADE DE LISBOA
FACULDADE DE MEDICINA DE LISBOA

UNIVERSIDADE
DE LISBOA

Copy number variation and Huntington's disease

Angelica Vittori

Ramo das Ciências Biomédicas
Especialidade – Neurociências

Outubro 2013

UNIVERSIDADE DE LISBOA
FACULDADE DE MEDICINA DE LISBOA

UNIVERSIDADE
DE LISBOA

Copy number variation and Huntington's disease

Candidata:
Angelica Vittori

Orientadores:
Prof . Doutor Tiago Fleming Outeiro
Doutor Flaviano Giorgini
Doutor Edward J. Hollox

Ramo das Ciências Biomédicas
Especialidade – Neurociências

A impressão são da exclusiva está dissertação foi aprovada pelo Conselho Cientifico da Faculdade de Medicina em reunião de 19 de Novembro de 2013.

Resumo

A variação de número de cópias (CNV em inglês) é uma modificação de uma sequência de DNA que apresenta uma inserção ou deleção em comparação com um genoma de referência com um número de cópias de N = 2. Com um comprimento variável, desde 50 pares de bases até várias megabases, as CNVs identificadas têm um tamanho médio de ~ 3 Kb e representam cerca de 4% do genoma humano. As CNVs, como outras variações genéticas, podem afetar directamente os níveis de expressão dos genes afectados. Os efeitos indirectos na expressão genética podem ser causados por alterações da posição, interrompendo o quadro de leitura do gene ou posteriormente, perturbando as redes de regulação genética. Foi demonstrado que as CNVs são em grande parte responsáveis pela evolução humana e diversidade genética entre os indivíduos. A relevância das CNVs no genoma humano foi salientada por vários estudos de associação que mostraram o efeito das CNVs na susceptibilidade a doenças neurodegenerativas, doenças de características complexas, e por serem a principal causa do aparecimento de doenças mendelianas ou por conferirem um fenótipo benigno.

A doença de Huntington (DH) é uma doença degenerativa e progressiva do cérebro que é fatal. Clínicamente é caracterizada por disfunção motora, distúrbios emocionais / psiquiátricos graves, declínio cognitivo e demência. A DH é monogénica e autossómica dominante, sendo causada por uma expansão anormal do trinucleótido citosina-adenina-guanina (CAG) no terminal 5' do gene da Huntingtina (*HTT*), localizado no cromossoma 4. Normalmente, a manifestação dos sintomas da doença dá-se por volta dos 40 anos, mas formas mais precoces (início <20 anos) e mais tardias (> 70 anos) também existem. O comprimento da expansão aberrante da repetição CAG no gene HTT está inversamente correlacionado com a idade de ocorrência (IO) da doença. O comprimento desta repetição explica ~ 50-70% na variação da IO sendo o restante da variação explicado por outros fatores genéticos e ambientais. Vários estudos têm como objectivo identificar modificadores genéticos da IO em DH, ou seja, nomeadamente um gene que quando

mutado tem a sua função alterada e determina o aparecimento do fenótipo associado com a mutação CAG. Há poucos modificadores genéticos identificados através de abordagens multi-translacionais de grandes estudos de coorte da doença para validação em vários modelos experimentais, representando, no entanto uma abordagem válida para identificar novos alvos terapêuticos para a DH.

O nosso objetivo neste estudo foi identificar genes candidatos com variação de número de cópias com potencial para modificar a IO na DH. Após a triagem de uma biblioteca de CNVs a partir de análise de dados matriz-CGH (disponível em linha através do navegador genoma UCSC), foram seleccionados vários candidatos: a região da β-defensine (*DEFB4*) e o transportador de glicose neuronal 3 (*SLC2A3*). Colocámos a hipótese de que a CNV na região de β-defensine (que varia de 2 a 8 cópias) poderia afectar o início da DH através de uma expressão variável do gene *DEFB4* (presente nesta região) que codifica para um péptido antimicrobiano importante. Um aumento do número de cópias, codificantes para um nível mais elevado de proteína, podem contribuir a inflamação neuronal presente na DH e consequentemente piorando a evolução da doença. Assim, investigámos o papel potencial da CNV do gene *SLC2A3* (que variam de 1 a 3 cópias), assumindo que o maior número de cópias pode ser protector na DH, codificando para um nível mais elevado do transportador de glicólise (GLUT3).

Analisámos a distribuição das CNVs dos nossos genes candidatos em indivíduos do estudo REGISTRY realizado pela Grupo Europeu da Doença de Huntington (EHDN), utilizando o Teste da Razão de Parálogos (PRT). PRT é uma tipagem de elevada capacidade de variação do número de cópias que se baseia em PCR quantitativo e relativo.

Determinámos o número de cópias da β-defensine em 490 indivíduos da nossa coorte de DH. A distribuição de CNVs de β-defensine nesta coorte não é significativamente diferente da população europeia sem DH (valor de p= 0,91) e também não está associada a uma alteração significativa da IO (valor de p = 0,41). Assim, podemos concluir que a CNV de β-defensine não é relevante na patologia da DH durante a fase inicial da progressão da doença. No entanto, com o nosso estudo, não podemos descartar qualquer papel potencial do gene *DEFB4* numa outra fase da doença.

Determinámos depois a distribuição de CNVs no gene *SLC2A3* em 988 indivíduos com a DH e não foi significativamente diferente da população britânica (valor de p =0.38). Curiosamente, verificou-se uma associação significativa entre a CNV de *SLC2A3* e uma variação de IO na nossa coorte de DH (valor de p = 0,028). Indivíduos com três cópias de *SLC2A3* mostraram um atraso do IO em comparação com o resto do grupo de até cerca de 6 anos. De modo a investigar a base molecular deste efeito protector da GLUT3 analisámos os níveis de expressão da proteína em linhas linfocíticas imortalizadas obtidas a partir de pacientes da nossa coorte, portadores de 1, 2 e 3 cópias de *SLC2A3* (5 por cada genótipo). Pela análise immunoblot detectou-se que um maior número de cópias no gene *SLC2A3* está associado a um maior nível de expressão da proteína (valor de p = 0,02). Assim, o efeito protector do aumento do número de cópias de *SLC2A3* é devido a uma maior expressão do transportador de glicose.

Para investigar a base funcional deste efeito protector, foi utilizado um modelo de *Drosophila* de DH para estudar o efeito de sobre-ou sub-expressão do funcional homólogo de *SLC2A3* em *Drosophila*, *Glut1*. Analisámos vários indicadores relevantes da doença, incluindo a neurodegeneração dos fotorreceptores, longevidade e taxa de eclosão. Descobrimos que em moscas com DH a sobreexpressão de *Glut1* pode reduzir significativamente a perda de rhabdomeres e portanto a neurodegeneração (valor de p <0,001). Pelo contrário, a redução ou a alteração génica deste transportador estão associados com um pior cenário que se manifesta por uma perda maior de rhabdomeres (valor de p = 0,00012 *), uma redução da longevidade média (valor de p = 0,034) e uma redução da taxa de eclosão (valor de p = 0,015). Alterações do principal transportador de glicose num modelo de *Drosophila* da DH agrava a progressão da doença numa fase inicial, e sobrerexpressão deste transportador protege da neurodegeneração na mosca adulta, confirmando os resultados na coorte de DH.

Do que sabemos, este é o primeiro estudo que investigou a função potencial das CNVs em DH. Conclui-se que a CNV no gene *SLC2A3* é um novo modificador genético do IO em grandes coortes de DH (988 amostras), em que un aumento de cópias do gene está associado a um efeito protector na DH. Testámos o efeito molecular da ocorrência desta CNV em linhas celulares obtidas directamente dos pacientes e descobrimos que o evento de

CNV do gene é responsável por um aumento da expressão de GLUT3 na DH, o que poderá repor a deficiência energética nos neurônios de pacientes com a DH e atrasar a neurodegeneração. Além disso, usámos um modelo de *Drosophila* de DH para testar a função funcional da homólogo de *SLC2A3* em moscas (*Glut1*), confirmando que uma sobre-expressão deste transportador pode diminuir a progressão da doença na mosca adulta, e a alteração do gene piora a progressão da doença num estadio inicial. É necessária futura investigação para reproduzir esta observação e confirmar o *SLC2A3* como um alvo de potencial terapeutico para o tratamento da DH.

** os dados referem-se a moscas no dia 1 pós eclosão.*

Palavras-chave: variação do número de cópia (CNV), doença de Huntington (DH), modificadores genéticos, DEFB4, SLC2A3, Drosophila.

Index

Index of the Figures

Index of the tables

Abbreviations

The gene or protein acronyms listed in this section refer to *Homo sapiens* genes and proteins except where specified.

18FDG 18F-fluorodeoxyglucose

^{1}H-MRS ^{1}H-magnetic resonance spectroscopy

AD Alzheimer's disease

ADEOAD autosomal dominant early onset AD

ADORA2A adenosine A2a receptor

ALS Amyotrophic lateral sclerosis

AMPK AMP-activated protein kinase

AMY1 salivary amylase gene

AO age of onset

APOE apolipoprotein E

APP amyloid-β protein precursor

array-CGH array-based comparative genomic hybridization

BAF B allele frequency

BB19 brain capillary endothelial cells

BBB blood-brain barrier

BDNF brain-derived neurotrophic factor

bp base pairs

C4 complement component C4

C4A complement component C4A

C4B complement component C4B

CAG cytosine-adenine-guanine

CBP CREB binding protein

CCR6 chemokine (C-C motif) receptor 6

CKB brain-type creatine kinase

CNS central nervous system

CNV copy number variation

CSF cerebral spinal fluid

DC dendritic cells

DEFB1 human β-defensin 1

DEFB103 human β-defensin 103

DEFB104 human β-defensin 104

DEFB105 human β-defensin 105

DEFB106 human β-defensin 106

DEFB107 human β-defensin 107

DEFB4 human β-defensin 2

DJ-1 parkinson protein 7

DOCK5 dedicator of cytokinesis 5

ECACC European Collection of Cell Cultures

EHDN European HD network

EMS ethyl methane sulfonate

EP enhancer–promoter

ER endoplasmatic reticulum

FCGR3B Fc fragment of IgG low affinity IIIb receptor

FcγR Fc γ receptors

FISH fluorescent in situ hybridization

FoSTeS fork stalling and a template switching

FPR formyl peptide receptor 1

GLUT glucose transporter

Glut1 glucose transporter 1 (Drosophila melanogaster)

GRIK2 glutamate receptor ionotropic kainate 2

GRIN2A glutamate receptor ionotropic N-methyl D-aspartate 2A

GRIN2B glutamate receptor ionotropic N-methyl D-aspartate 2B

HAP-1 huntingtin-associated protein 1

HD Huntington's disease

Hdh huntingtin (*Mus musculus*)

HEAT Huntingtin Elongator factor3 the regulatory A subunit of protein phosphatase 2A and TOR1

HGDP-CEPH Human Genome Diversity Cell Line

HR homologous recombination

HTT huntingtin

IFN-γ interferon-gamma

IGF-1 insulin growth factor 1

IL-1β interleukin-1β

IL-α interleukin-1α

indel insertion-deletion

kb kilo base

KLK6 kallikrein-related peptidase 6

KMO kynurenine 3-monooxygenase

LCLs immortalized lymphocyte cell lines

LCR low-copy repeats

LD low dNTPs

LPS lipopolysaccharide

MAF minor allele frequency

MAPH multiplex amplifiable probe hybridisation

MAP-kinase mitogen-activated protein kinase

MAPT microtubule-associated protein tau gene

Mb megabases

MEOX2 mesenchyme homeobox 2

MHC major histocompatibility complex

ML Maximum Likelihood

MLPA multiplex ligation-dependent probe amplification

MMEJ microhomology mediated end joining

MRI magnetic resonance imaging

MRI magnetic resonance imaging

mTOR mechanistic target of rapamycin serine/threonine kinase

NAHR non allelic homologous recombination

NCoR nuclear co-repressor

NF-κB nuclear transcription factor κB

NGS next generation sequencing

NHEJ non homologous end joining

NMDA acid N-metil-D-aspartic

NRSF neuro-restrictive silencer factor

p53 tumor protein p53

PARKIN parkin

PCR polymerase chain reaction

PD Parkinson's disease

PET positron emission tomography

PET positron emission tomography

PFGE pulsed field gel electrophoresis

PGC-1α peroxisome proliferative activated receptor gamma coactivator 1 alpha

PINK1 PTEN induced putative kinase 1

PKC protein kinase C

PMP22 peripheral myelin protein 22

polyQ polyglutamine

PRT paralogue ratio test

PS1 presenilin 1

PS2 presenilin 2

qPCR real-time quantitative PCR

Rab11 CG5771 gene product from transcript CG5771-RA (*Drosophila melanogaster*)

RAI1 retinoic acid induced 1

r-BD1 defensin beta 1 (*Rattus norvegicus*)

r-BD2 defensin beta 2 (*Rattus norvegicus*)

REDVR restriction enzymatic digest variant

REST repressor element-1 silencing transcription factor

RNAi RNA interference

RpL32 ribosomal protein L32 (*Drosophila melanogaster*)

SLC2A14 solute carrier family 2 (facilitated glucose transporter), member 14

SLC2A3 solute carrier family 2 (facilitated glucose transporter), member 3

Slc2a3 solute carrier family 2 (facilitated glucose transporter), member 3 (*Mus musculus*)

SLC30A3 solute carrier family 30 zinc transporter member 3

SLE systemic lupus erythematosus

SMER small-molecule enhancer of rapamycin

SMN1 survival motor neuron 1

SNCA α-synuclein

SNPs single-nucleotide polymorphisms

SOD1 superoxide dismutase-1

Sp1 specificity protein 1

SV structural variation

TBP TATA-binding protein

TBZ Tetrabenazine

TLRs toll-like receptors

TNF-α tumor necrosis factor-α

UAS upstream activating sequence

UCH-L1 ubiquitin carboxyl-terminal esterase L1 ubiquitin thiolesterase

UGT2B17 UDP-glucuronosyltransferase 2B17

UPS ubiquitin-proteasome system

USP32 ubiquitin specific peptidase 3

Abstract

Huntington's disease (HD) is a fatal neurodegenerative disorder caused by the expansion of an unstable triplet repeat within the huntingtin gene. The length of this repeat is inversely correlated with the age of onset (AO) of the disease, which ranges from 1 to 80 years of age. The length of this repeat explains 50-70% of the variance of AO with the remaining variation attributable to environmental and other genetic factors.

Copy number variation (CNV) is a structural variation of the human genome wherein a genomic sequence is duplicated or deleted compared to a reference genome. As CNVs have the potential to affect gene expression either directly by dosage effects or indirectly by affecting gene product interactions and pleiotropic effects, they are found to be associated with phenotypic variance, disease susceptibility and Mendelian disorders.

The aim of the study was to investigate CNVs as candidate genetic modifiers of the AO in HD. Specifically we investigated CNV of the human β-defensin region (including *DEFB4*) and CNV involving *SLC2A3*, with potential impacts, respectively, in the neuroinflammatory response and in neuronal glucose uptake. CNVs were analysed within a large HD sample cohort (provided by EHDN) using the paralogue ratio test (PRT) to test their potential impact on a variance of the AO in HD.

In 490 HD individuals analysed the frequency distribution of β-defensin copy number was shown to be equivalent to the general European population. Furthermore, no significant association was shown between β-defensin CNV and a variance of the AO in HD.

987 HD patients were genotyped for *SLC2A3* CNV and a modest but significant association with a variance of AO in HD was found (p value = 0.028). Individuals with three copies showed a delay in the AO of up to nearly 6 years compared to individuals with one or two copies. In order to test if *SLC2A3* CNV affects gene expression 15 cell lines from patients with different *SLC2A3* copy number were immunoblotted for GLUT3 (encoded by *SLC2A3*) and it was found that the protein level was significantly correlated with the genomic copy number (p value = 0.020). Therefore, we concluded that *SLC2A3* CNV is a genetic modifier of

the AO in HD, associated to a variance of the disease onset and affecting the gene expression.

To investigate the functional basis of this effect, we analysed lines showing over- or under-expression of the functional homolog of *SLC2A3* (*Glut1*) in a *Drosophila* model of HD. After analysing several disease-relevant metrics, including neurodegeneration of the photoreceptors, eclosion rate and longevity, we found that gain and loss of *Glut1* expression can delay and worsen, respectively, neurodegeneration in HD flies.

In conclusion, β-defensin genomic copy number is not associated with modulation of HD AO. On the other hand, *SLC2A3* CNV is a genetic modifier of the AO in HD, and likely has functional consequences, based on our findings in patient cells and in a *Drosophila* HD model.

The study reported in this thesis has resulted in the publication of a paper entitled:

"β-Defensin Genomic Copy Number Does Not Influence the Age of Onset in Huntington's Disease"

Angelica Vittori, Michael Orth, Raymund A.C. Roos, Tiago F. Outeiro, Flaviano Giorgini, Edward J. Hollox, and REGISTRY investigators of the European Huntington's Disease Network

published on Journal of Huntington's Disease 2 (2013) 107–124,

and the preparation of another paper submitted to Human Molecular Genetics:

"Copy number variation of the neuronal glucose transporter gene SLC2A3 modifies the age of onset in Huntington's disease"

Angelica Vittori, Carlo Breda, Mariaelena Repici, Michael Orth, Raymund A.C. Roos, Tiago F. Outeiro, Flaviano Giorgini, Edward J. Hollox, and REGISTRY investigators of the European Huntington's Disease Network

CHAPTER ONE: INTRODUCTION

1.1 Copy Number Variation

Variation in the human genome encompasses single-nucleotide polymorphisms (SNPs), small insertion-deletion (indel) polymorphisms, variable numbers of repetitive sequences, and genomic structural variations (Iafrate *et al*. 2004). Structural variation (SV) is a broad term that refers to genetic variants bigger than ~1 kilo base (kb), able to modify the chromosomal architecture (Feuk, Carson, Scherer 2006); it includes balanced changes, as inversions and reciprocal translocations, and alteration of DNA copy number, namely copy number variation (CNV) (Hurles, Dermitzakis, Tyler-Smith 2008). In the last six decades, some SV of phenotypic relevance have been catalogued through the use of molecular genetics and cytogenetic tools. Only with the advent of the array-based comparative genomic hybridization (array-CGH) technology coupled with the publication of the draft human sequence in 2001 was it possible to analyse SV in large scale data sets, of phenotypically normal and diseased populations. Through use of array-CGH technology, several studies identified the presence of a number of CNVs, highlighting their role as the largest component of the SV within the human genome (Conrad *et al*. 2010; Iafrate *et al*. 2004; Kidd *et al*. 2008; Redon *et al*. 2006; Sebat *et al*. 2004; Tuzun *et al*. 2005).

CNV is an alteration of a DNA sequence showing an insertion or deletion compared to a reference genome with a copy number of N = 2 (Feuk, Carson, Scherer 2006). The extent of detected CNVs ranges from 50 base pairs (bp), as an operational demarcation from indels, up to megabases (Mb), with a median size of ~ 3 kb (Alkan, Coe, Eichler 2011; Conrad *et al*. 2010). CNV can arise in a diallelic or multiallelic configuration (Wain, Armour, Tobin 2009): diallelic variation takes place with homozygous or heterozygous deletion or duplication of a DNA fragment; a multiallelic CNV consists of homozygous or heterozygous deletion or duplication as equal or/and unequal change in each chromosome (Figure 1). Moreover, CNV can manifest as a complex rearrangement of a DNA region, where different regions, present in a variable number of copies, are involved in the same event (Figure 1).

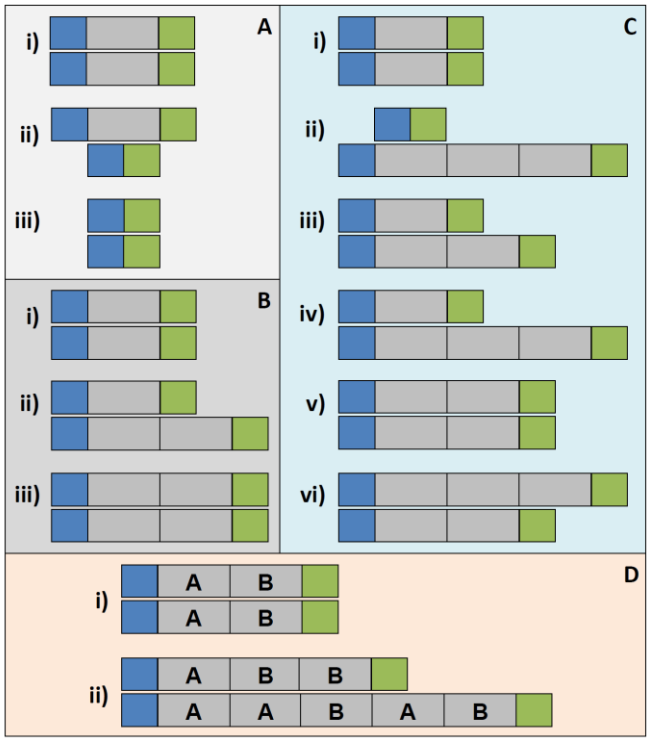


Figure 1. Diallelic and multiallelic configurations of CNV.

The copy number variable site (grey), flanked by invariant regions (blue and green), is involved in a diallelic (A) deletion and (B) duplication, each showing the site with (i) normal diploid set, (ii) heterozygous modification, and (iii) homozygous modification. (C) Multiallelic locus showing (i) normal copy number, (ii) multiple rounds of duplication on one chromosome and a deletion on the homologous chromosome, (iii) duplication on one chromosome and no deletion on the homologous chromosome, (iv) multiple rounds of duplication on one chromosome and no deletion on the homologous chromosome, (v) one round of duplication on each chromosome, (vi) one round of duplication on one chromosome and multiple rounds of duplication on the homologous chromosome. (D) Multiallelic locus showing (i) normal copy number and (ii) a complex rearrangement of the two different alleles A and B. [Adapted from Wain, Armour, Tobin 2009].

Despite the increasing number of studies investigating CNV, it is not possible to determine an exact coverage of CNV in the human genome. Due to limitations in the resolution of CNV detection tools, many observed CNVs are described with an approximate size and the location of the variant boundaries is overestimated (Hurles, Dermitzakis, Tyler-Smith 2008). Furthermore, with the absence of standardised controls, the variability between array-based platforms and the different calling algorithms used for the discovery of CNVs in several independent studies (Pinto *et al*. 2011) there has certainly been a large

number of redundant entries of these variants (> 610,834), as the Database of Genomic Variants reports (http://dgvbeta.tcag.ca/dgv/app/home?ref=NCBI36/hg18).

A recent study performed by the Genome Structural Variation consortium (Conrad *et al*., 2010), based on tiling oligo array-CGH platform, analysed CNV events of 450 individuals from the HapMap project (International HapMap Consortium 2005) and 45 individuals from China; the identified CNVs (8,599) extend for 3.7% of the human genome, coverage that realistically fits previous analyses, where the total amount of duplicated sequence in the human genome is nearly 5% of the entire genome (Bailey *et al*. 2002; Cheung *et al*. 2003; Cheung *et al*. 2001; She *et al*. 2004). In particular, ~ 18% of the validated CNVs overlap with coding regions of the human genome and an average of 45 OMIM (Online Mendelian Inheritance in Man, http://www.omim.org/) genes per analysed sample were showed to be affected by 48 CNVs, highlighting the role of these variants as a relevant factor in human diversity, evolution and disease susceptibility (Conrad *et al*. 2010).

1.1.1 Mechanisms of CNV formation

CNV occurs as *de-novo* mutation in germline and in somatic cells at a significant rate, and can become heritable polymorphisms. The CNV mutation rate has been estimated to range from 1.7×10^{-6} to 1.0×10^{-4} per locus per generation and can vary broadly at different loci and between different tissue, showing somatic mosaicism (Breckpot *et al*. 2011; Bruder *et al*. 2008; Lupski 2007; Notini, Craig, White 2008; Piotrowski *et al*. 2008; van Ommen 2005).

CNV formation is the result of changes in the chromosomal structure, changes which themselves arise when two formerly separated DNA fragments become juxtaposed (Hastings *et al*. 2009). CNV is more likely to occur in the proximity of centromeres and telomeres (Guryev *et al*. 2008; Nguyen, Webber, Ponting 2006) and in regions containing highly homologous duplicated sequence and more specifically segmental duplication (also called “low-copy repeats” (LCR)) (Sharp *et al*. 2005), which themselves are blocks that exhibit a high degree of nucleotide sequence identity (commonly > 95%) and enclose large genomic distances (1-100kb) (Eichler 2001). Segmental duplications are normally located in

pericentromeric and subtelomeric regions (Eichler 2001), usually associated with chromosomal instability or evolutionary rearrangement (Sharp *et al*. 2005), suggesting the implication of segmental duplications with genomic disorders (Sharp *et al*. 2006; Stankiewicz and Lupski 2002). Through ectopic homologous recombination (HR), specifically non allelic homologous recombination (NAHR), LCRs participate in the formation of recurrent CNV.

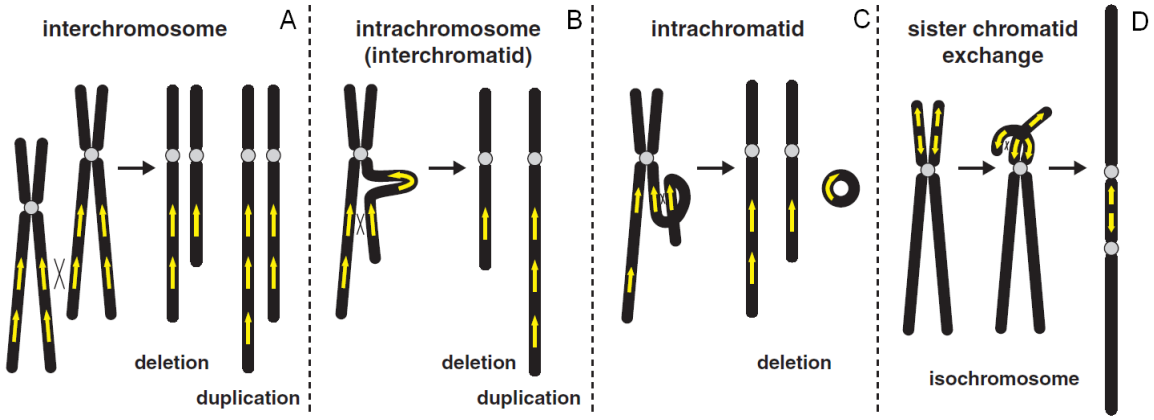


Figure 2. NAHR mechanisms.

(A,B) Deletion or duplication can occur by NAHR in two ways, interchromosomal crossover and interchromatidal crossover. (C) Deletion can also arise by intrachromatidal crossover. (D) NAHR between inverted LCRs on sister chromatids can also result in isochromosome formation (normally observed in somatic cells within a tumour). [Adapeted from Liu at al. 2012].

NAHR is a mechanism that leads to *de novo* mutation events in germline and somatic cells, and is a major cause of copy number change occurring when control of allelic recombination fails (Liu *et al*. 2012). The likelihood of NAHR event occurring is proportional to the LCR or paralogous repeat sequence length and is inversely correlated to the distance between the loci involved in the potential recombination event (Liu *et al*. 2011). Ectopic recombination produces change in copy number, which results in a deletion and reciprocal duplication. NAHR events can follow different model or genomic rearrangements involving paralogous LCRs on both single chromosome and homologous chromosomes, or inverted LCRs on sister chromatids (Figure 2) (Lange *et al*. 2009; Liu *et al*. 2012). Furthermore, NAHR

can occur between LCRs in non-homologous chromosome, recurrent translocations may ensue (Ou *et al*. 2011).

Less frequent CNV can originate from mechanisms of DNA repair that, with lower stringency, use very limited homology (2-15 bp) or no homology. The breakpoint regions of these CNVs are characterized by blunt ends, microhomologies, and small insertions, suggesting the involvement of a nonhomologous repair mechanism in their formation (Lee, Carvalho, Lupski 2007; Luo *et al*. 2011b; Shaw and Lupski 2005).

Non homologous end joining (NHEJ) is with HR one of the main pathways for DNA double-strand break repair in eukaryotic cells; it is a stable but imprecise mechanism and compared to HR it does not require substrates with extended homology (Lieber *et al*. 2003; Lieber 2008). Through rearrangements within scattered breakpoints, NHEJ can lead to translocations and telomere fusion (Espejel *et al*. 2002) as a consequence of erroneous loss or addition of several nucleotides at the end joining point (Lieber 2008). Several studies showed the role of NHEJ in the formation of deletion CNVs (Luo *et al*. 2011a; Shaw and Lupski 2005; Toffolatti *et al*. 2002) and, combined with HR, of duplication CNVs (Inoue *et al*. 2002; Lee, Carvalho, Lupski 2007).

An alternative repair mechanism to NHEJ is microhomology mediated end joining (MMEJ) which requires different intermediary factors compared to NHEJ and results in a deletion with microhomology at the repair junction. MMEJ is involved in genomic rearrangements in spontaneous and therapy related cancer (Bennardo *et al*. 2008).

Non-recurrent CNVs showing highly complex structure, that cannot be readily explained by NHEJ, are attributable to fork stalling and a template switching (FoSTeS) mechanism (Lee, Carvalho, Lupski 2007). FoSTeS is a replication based model able to generate large genomic duplications of several mega base, which may lead to gene duplications/ triplication or exon shuffling (Bi *et al*. 2009; Zhang *et al*. 2009). During DNA replication, the active replication fork can stall and switch between complementary templates, sharing microhomology and apart in the genome, to anneal and prime DNA replication (Kitamura, Blow, Tanaka 2006; Lee, Carvalho, Lupski 2007; Zhang *et al*. 2009). FoSTes as a replicative method occurs in mitosis (Lee, Carvalho, Lupski 2007) or can be induced in stressed cells (Arlt *et al*. 2009;

Slack *et al*. 2006). In case of break-induced replication, CNV may also occur by way of a repair model, suggested by Hastings *et al*. called microhomology-mediated break-induced replication, which enables formation of complex structure by several round of template switching (Hastings, Ira, Lupski 2009).

1.1.2 Methods of CNV identification and genotyping

Historically, only large chromosomic rearrangements were detected in the context of severe developmental disorders (such as Down syndromes) by cytogenetic analysis (Jacobs *et al*. 1959) and were not referred to in any systematic fashion. Only the advance of genomic investigative tools, such a microarray and sequencing technologies, applied to genome-wide studies of apparently healthy individuals made possible the discovery and genotyping of a broad range of CNVs. Several techniques are currently available for the detection and genotyping of CNVs (Table 1).

1.1.2.1 Microarray-based techniques

Microarrays, the experimental workhorse of CNV discovery and genotyping (Alkan, Coe, Eichler 2011), are represented primarily in the context of CNV studies by array CGH (Pinkel *et al*. 1998) and SNP microarrays (Huang *et al*. 2004) and both platforms infer the copy number by comparison with a reference genome, which could be a single sample or population data set (Figure 3).

The array CGH technology (Kallioniemi *et al*. 1992) is based on the hybridization of two different labelled samples, a test and a reference, to a set of specific targets, typically oligonucleotides (Conrad *et al*. 2010; Kidd *et al*. 2008) or, historically, bacterial artificial chromosome clones (Redon *et al*. 2006). The signal ratio between test and reference is normalized and converted to a $\log_2$ ratio: an increase of the $\log_2$ ratio indicates a gain of copy number of the test region compared to the reference; conversely, a diminution represents a decrease in copy number (Figure 3) (Oostlander, Meijer, Ylstra 2004; Pinkel *et al*. 1998). It is of crucial importance that an appropriate reference genome(s) is chosen to minimise the risk of incorrect CNV calling. For example, if a non-detected deletion is present

Table 1. Comparative table between CNV detection and genotyping techniques.

	array-CGH	SNP Array	Fiber FISH	PFGE	NGS	MAPH	MLPA	PRT
Detection	Relative Copy Number	Relative Copy Number	Absolute Copy Number	Inferred Absolute Copy Number	Absolute Copy Number	Relative Copy Number	Relative Copy Number	Relative Copy Number
Sample	2-5 µg DNA	2-5 µg DNA	Cells	2-5 µg DNA	2-5 µg DNA	0.5-1 µg DNA	100-200 ng DNA	5 - 10 ng DNA
Loci	≥ 2 Million	≥ 2 Million	Single	Single	Genome-Wide	Up To 40	Up To 50	Single
Throughput	High	High	Low	Low	High/moderate	High	High	High
Minimum Resolution	5 – 10 kb	5 – 10 kb	≥ 1 kb	0.5 - 1 kb	≥ 1 kb	100 bp	100 bp	100 bp
Cost Per Sample	Moderate	Moderate	Low	Low	High	Low	Low	Low
Time	≥ 24 h	≥ 24 h	≥ 24 h	2- 3 days	2 -3 days	≥ 24 h	≥ 24 h	4 h
Labor Requirement	Moderate	Moderate	High	High	High	Low	Low	Low

[Adapted from Cantsilieris, Baird, White 2012].

in the reference genome then a duplication event would be called in all the analysed samples.

SNP arrays generate a similar output by comparing the signal intensity for a sample compared to a collection of the normalized total intensity of reference hybridizations (Log R Ratio). Furthermore, the B allele frequency (BAF), namely the normalized measure of the allelic intensity ratio of two alleles of each SNP studied, is analysed. The combination of the Log R Ratio and BAF ratio can be used to infer copy number changes in the genome especially for multi-allelic copy number (Figure 3) (Wang and Bucan 2008). The BAF metric is not informative in the case of homozygous or balanced structural variants.

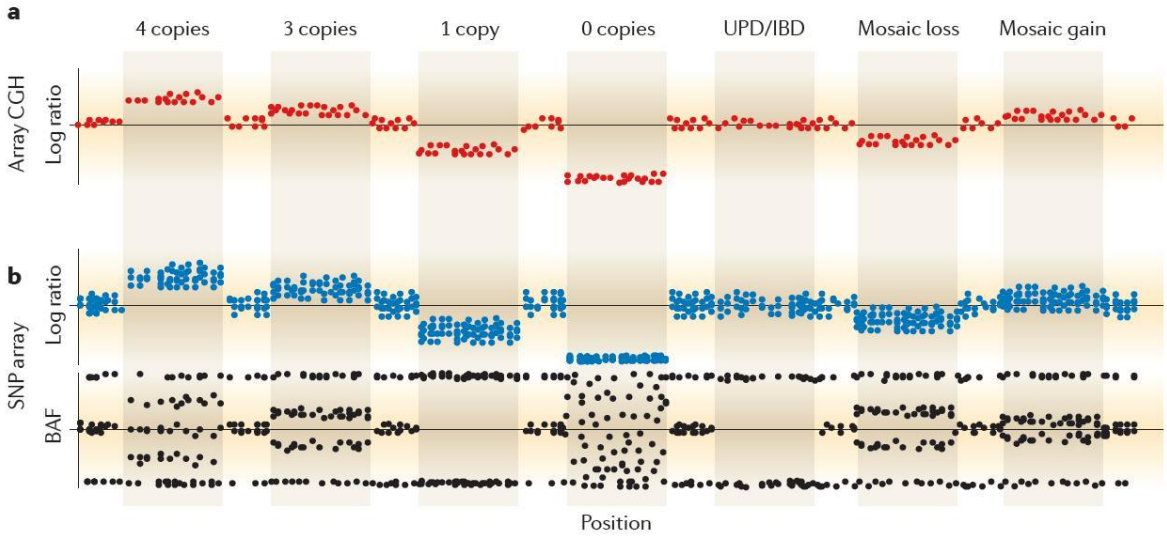


Figure 3. Analysis of intensity signal from array CGH and SNP array.

Array CGH (a) is informative on CNV with reduced signal noise compared to SNP array (b), that through BAF is informative in case of multi-allelic-copy number: in the case of AA, AB and BB genotype the BAF ratio should be 0, 0.5 and 1.0 respectively; for higher copy number the BAF resulting ratios are indicative of the allelic asset, e.g. in case of ABB genotype the ratio would be 0.33/ 0.67. [Adapted from Alkan, Coe, Eichler 2011].

Microarray technology proffers a tangible advantage in terms of throughput and cost allowing one to assay CNV architecture for large data sets, an essential requisite in the identification of rare-disease variants. Disadvantages of the microarray technology are the detection of copy number differences of sequences that are not present in the reference genome used for the probe design (Kidd *et al*. 2010) and in the localization of breakpoints at

a single-base pair level. Microarray technology is not yet suitable for the analysis of repeat-rich and duplicated regions where the probe coverage would not be accurate and the CNV call is not sensitive enough. Furthermore, the algorithm applied for the CNV call accounts for another important limitation: the analysis of the same raw data with different algorithms leads to a higher variability in the CNV calling than when the same algorithm is used on raw data from identical samples, samples analysed by different laboratories (Pinto *et al*. 2011).

1.1.2.2 Hybridization-based techniques

Southern blotting is a hybridization technique used for typing structural rearrangements. Southern blotting procedure consists of fragmentation of DNA with a cutting restriction enzyme, separation of the products in electrophoretic run and transfer to a nylon membrane (blotting), which will be treated with a labelled DNA probe for the identification of the target band (Southern 1975). Through analyses of band intensities and normalization on a copy stable locus band, Southern blotting can be used for typing CNV genes (Aitman *et al*. 2006). However, several disadvantages of this method such as the intense work flow, the high DNA amount required and the scarce accuracy for the distinction of consequential copy number make the Southern blotting alone not suitable for CNV studies.

Pulsed field gel electrophoresis (PFGE) is an umbrella term referring to technique based on the digestion of DNA with a rare-cutter restriction enzyme, followed by separation on a pulse field electrophoresis gel and Southern-blot analysis (Schwartz and Cantor 1984). PFGE can accurately resolve CNV thanks to the separation of restriction fragments, which size is specific for each copy number, on a periodic alternate electric filed gel. PFGE is used to type complex CNVs and to infer absolute copy number (Aldred, Hollox, Armour 2005; Yang *et al*. 2007).

Fiber FISH (fluorescent in situ hybridization) is an accurate technique based on the hybridization of a fluorescent labeled probe to microscope slide within DNA in metaphase or interphase and allows the identification of complex chromosomal rearrangements (Price 1993). Compared to others techniques, fiber FISH allows to identify high copy number (> 10) and is the only one that specifically consents to call CNV per allele, which is important for

studies of inheritance and disease (Perry *et al*. 2007; Pinkel *et al*. 1998; Sebat *et al*. 2004; Trask *et al*. 1998).

Due to their limited throughput and high laboriousness, PFGE and fiber FISH have a restricted application. They are commonly used as experimental validation methods and FISH is specifically applied to clinical test for chromosomal abnormalities screening.

1.1.2.3 Next generation sequencing techniques

Next generation sequencing (NGS) is a term that refers to a wide range of non-Sanger sequencing techniques based on a computational high throughput approach (Korbel *et al*. 2007; Tuzun *et al*. 2005; Volik *et al*. 2003). In NGS, the sequence of interest is processed as a single molecule template or PCR-clonal colonies, both are tethered to a solid surface or support and sequenced and analysed in parallel with the application of specific algorithms, reducing costs and time of the procedure (Medvedev, Stanciu, Brudno 2009; Metzker 2009). Compared to hybridization techniques, NGS technology is theoretically suitable for discovering all genomic variants with a much higher precision. CNV can be detected and genotyped using different NGS strategies, which differ for sensitivity and specificity depending on the dimension and type of CNV: read-depth, read-pair, split-read and *de novo* assembly methods (Alkan, Coe, Eichler 2011).

The read pair strategy can be used for detecting any sort of structural variations, using paired-end reads which are assessed for span and orientation compared to a reference genome. For example, a deletion or duplication are detected when an alteration of the distance between the read pairs occurs in the reference genome and a new insertion when one of the read does not map the reference genome (Kidd *et al*. 2008; Korbel *et al*. 2009; Tuzun *et al*. 2005).

The high coverage of NGS makes it possible to infer absolute copy number of known targets based on read-depth strategy. The read-depth approach infers CNV of a genomic locus detecting lower or higher than expected sequencing coverage of that region, assuming that the coverage of any region follows a Poisson distribution (or modified Poisson model). The read depth approach can also be used to discriminate among paralogs within a highly duplicated gene family (Alkan *et al*. 2009; Sudmant *et al*. 2010).

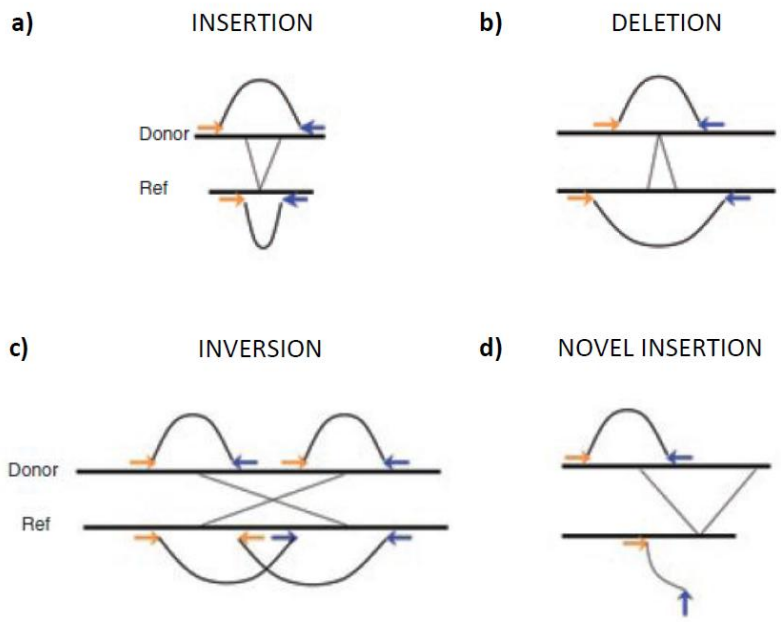


Figure 4. Read pair signatures.

The read pair is represented by the two arrows, sampled from the donor and mapped on the reference genome. (a, b) Insertion and duplication are detected respectively by a reduction or increased of the mapped distance between the two reads. (c) Inversion can be detected as a well from a variation of the order of the mapped reads. (d) New insertion can be detected when only one of the reads is mapped in the reference genome. [Adapted from Medvedev, Stanciu, Brudno 2009].

A split read approach can be used to define the breakpoint of a structural variant detecting deletions and small insertions with a single-base-pair resolution, respectively analysing reads misalignments in the test or reference genome. Split read methods have a limited application in CNV detection due to the use of longer reads than the other NGS methods and have less power in highly repeated regions (Abyzov and Gerstein 2011; Ye *et al*. 2009).

De novo genome assembly using NGS technology is the most accurate strategy for the detection of novel CNV and any sort of genomic variation. This approach has been applied to a small number of samples highlighting 4.1 terabases of raw sequence not present in the reference genome and several novel SVs (Altshuler *et al*. 2010; Mills *et al*. 2011; Wheeler *et al*. 2008). Optimisation of NGS technologies, reducing the laboriousness

and improving the algorithms used is necessary to adapt this method for large scale studies, which is a basic requirement for studying phenotypic effect of the detected variants.

NGS technologies allow the identification of break points, inversions, novel insertions, and specifically through the read-depth analysis is possible to distinguish between paralogous copies of duplicate genes within a family and to predict the absolute copy number call of genomic duplication/deletion intervals (Alkan *et al*. 2009). The combination of two complementary signatures for detection of structural variants by integrating read pair and read depth or split read approaches is a more accurate and efficient strategy than using a single method (Mills *et al*. 2011; Rausch *et al*. 2012). Nevertheless, there are several limitations in the NGS technologies: the algorithms used and the scarce manageability of large data-sets produced make the analyses time expenditure quite high; moreover, technical issues related to the length of the read analysed, which are too short for efficient assembly. These drawbacks make these techniques not yet suitable for large scale studies, reducing their power.

1.1.2.4 PCR based methods

The discovery of new SV by microarray or sequencing techniques requires the application of high narrow thresholds to restrain the detection of false positives. Genotyping analyses, designed for investigation of a few loci, are powerful tools characterized by less stringency, high sensitivity and specificity. CGH array, SNPs array and NGS methods can be applied and customized for CNV genotypes, presenting similar advantages and limitations. PCR (polymerase chain reaction)-based methods are single base pair sensitive tools that allow one to pinpoint breakpoint mutation locus and quantify copy number change; once optimised they permit analyses of large data set with low-cost and time saving frame.

Real-time quantitative PCR

Real-time quantitative PCR (qPCR) is a PCR-based method that allows absolute or relative quantification of mRNA levels and small genetic regions. The copy number is inferred as a ratio between the amplification products of a gene of interest versus a reference gene, amplicons that are detected at several time points by fluorescent signals of the incorporated probes. The kinetics of amplification of the two targets are required to be similar and equally efficient in order to make sure that the detected product will be an

accurate indication of abundance of the initial DNA product (Gonzalez *et al*. 2005; Higuchi *et al*. 1993). Although qPCR is characterized by high-throughput and simple work-flow, different studies, through comparison of with other PCR-based techniques for CNV calling, showed that qPCR is not suitable for complex CNV genotyping as it is inaccurate in the distinction of small fold changes, causing extensive overlap between copy number integer classes and leading to biased CNV associations (Aldhous *et al*. 2010; Field *et al*. 2009; Fode *et al*. 2011).

Multiplex amplifiable probe methods

Multiplex amplifiable probe hybridisation (MAPH) and multiplex ligation-dependent probe amplification (MLPA) are high-throughput techniques for the relative quantification of copy number changes of genomic DNA, using labelled probes for the simultaneous detection of several loci (up to ~ 45) (Armour *et al*. 2000; Schouten *et al*. 2002). These techniques are single base pair specific and allow analysis of large number of samples with reduced costs. The copy number of a specific target is inferred by the relative intensity between the amplification products, which can be distinguished from other targets thanks to a target-specific stuffer bound to one of the oligos, besides the primer used for the PCR reaction (Figure 5).

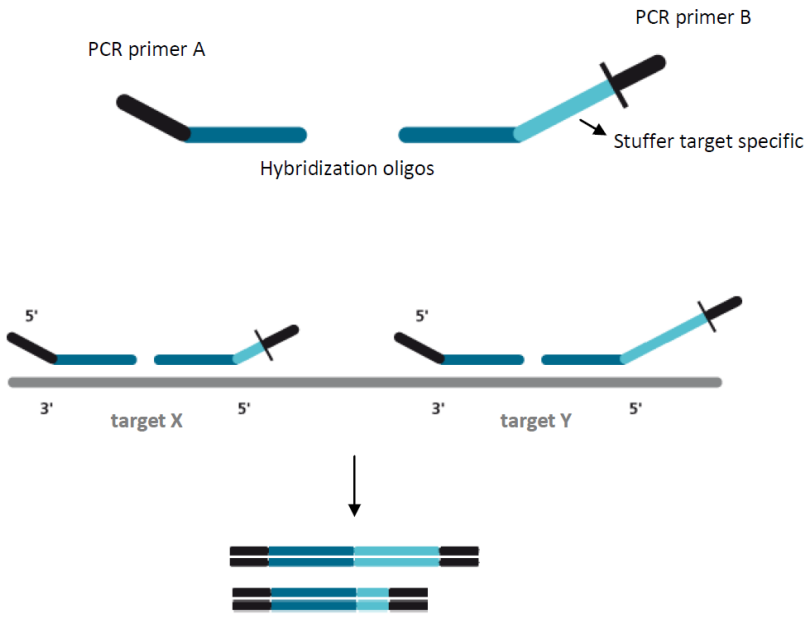


Figure 5. Basic principle of MAPH and MLPA techniques.

The amplification products are analysed according to the label used: polyacrylamide gel electrophoresis for MAPH (radioactive label), capillary electrophoresis for MLPA (fluorescent label). With the detection of events located on a single DNA fragment, this technology allows analysis of rare subpopulations or individual CNV events. Disadvantages reside in the laborious optimisation of these techniques and in the possibility to detect false mosaicism resulting from a lack of alignment of the primers caused by the presence of polymorphisms.

Paralogue Ratio Test

The paralogue ratio test (PRT) is a high-throughput typing of copy number variation method based on a relative quantitative PCR (Armour *et al*. 2007). The method requires small amounts of genomic DNA (5-10 ng) and consists of a PCR reaction performed under quantitative conditions (30 cycles), followed by capillary electrophoresis. PRT employs a single pair of fluorescent labelled primers designed to amplify simultaneously a fragment within a variable repeat locus of interest and within a reference locus that is not copy number variable (Figure 6). The products are resolved by size difference (optimally designed to range between 2 to 50 bp in order to minimize amplification variance) in capillary electrophoresis run.

The copy number is inferred from the ratio between the test and the reference locus amplification products. For complex CNV regions multiple PRTs can be performed in a single PCR reaction and analysed in parallel; PRT ratios, normalized for control copy number samples, are pooled to infer the copy number by likelihood calculations (Hollox, Armour, Barber 2003; Hollox, Detering, Dehnugara 2009; Walker, Janyakhantikul, Armour 2009).

One of the main challenges of this method is the design of the primers, which must anneal only the reference locus and copy number variable locus. This can be achieved with the help of an algorithm, which can quickly design couple of primers, suitable for PRT methods. The algorithm blasts the region of interest with the entire genome sequence, masking repeated regions, in order to find specific and unique paralogous regions and in combination with primer design software, selects the oligos annealing only for those (Veal *et al*. 2013).

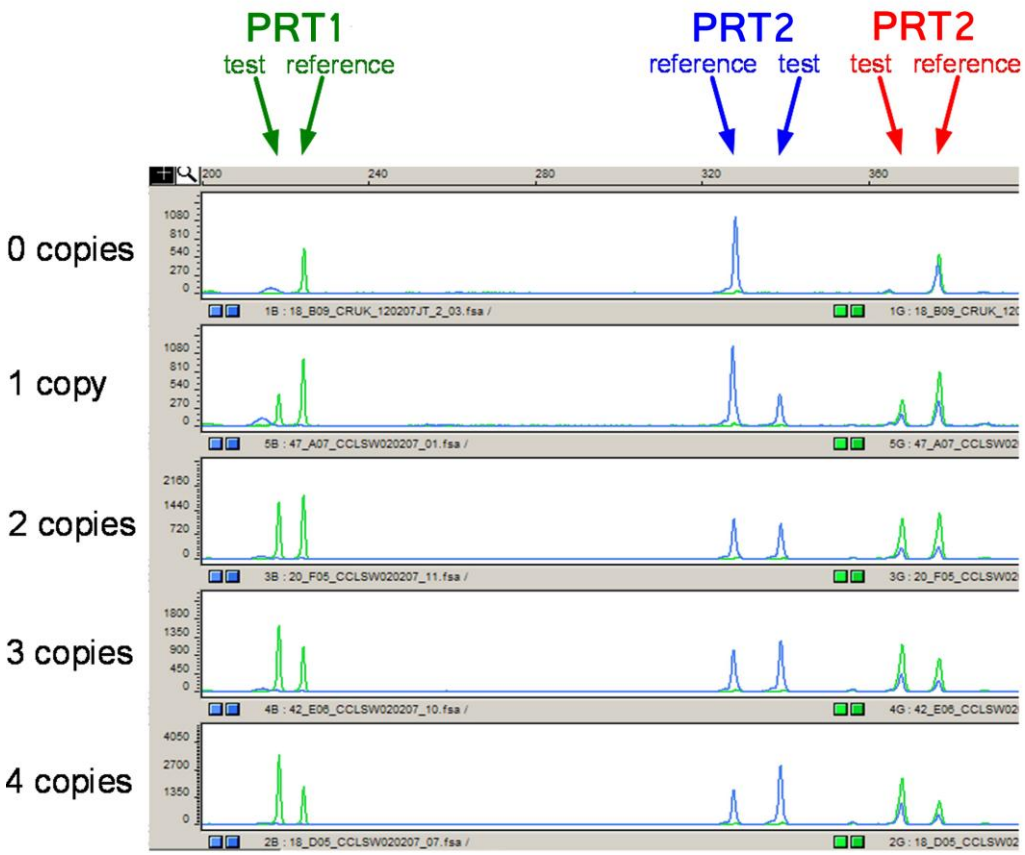


Figure 6. Capillary electropherogram of a multiplex PRT analysis.

Each electropherogram shows the amplification product of three different PRT (in duplicated for PRT3, using different labelled probes) where the ratio between test and reference peak areas is a clear indication of the variable copy number of the test compared to the reference locus. [Adapted from Walker, Janyakhantikul, Armour 2009].

A disadvantage of this test is the low accuracy in calling among odd and even copy number, this however can be resolved by combining the PRT to other assay such as a restriction enzymatic digest variant (REDVR) or an indel assay (Abu Bakar, Hollox, Armour 2009; Aldred, Hollox, Armour 2005). These methods reduce disambiguation between CNV calls based on the allelic distribution of SNPs or indel polymorphisms between the several units of the repeated region (Figure 7).

A REDVR method consists in the amplification of a SNP (commonly heterozygous in normal population) within the copy number variable locus followed by SNP-allele specific-restriction enzyme digestion. The digestion products are resolved by capillary electrophoresis. In the case of heterozygous genotype, the ratio between the digested and

undigested fragments area informs the distribution of two alleles that must agree with the number of copy of that region: for instance a ratio of 5:3 can only fit a 8 copy number model (in case of a homozygous genotype there is a single product which is not informative). Redundant ratios from multiple numbers of copies (2:2, 3:3 from 4 and 6 copies number or 4:2, 6:3 from 6 and 9 copies number) does not allow the use of this method alone, which is, in fact, coupled with the PRT approach to distinguish between consequential copy number. The indel assay is based on a similar rationale, where an indel within the copy number variable region is amplified instead of SNP. The PCR products are directly analysed on capillary electrophoresis and, similarly for the REDVR assay, the ratio between the different products is informative of the copy number (Figure 7). Data from the REDVR assay and indel assays are pooled with PRT data in maximum likelihood calculations to infer copy number calls (Abu Bakar, Hollox, Armour 2009; Aldred, Hollox, Armour 2005).

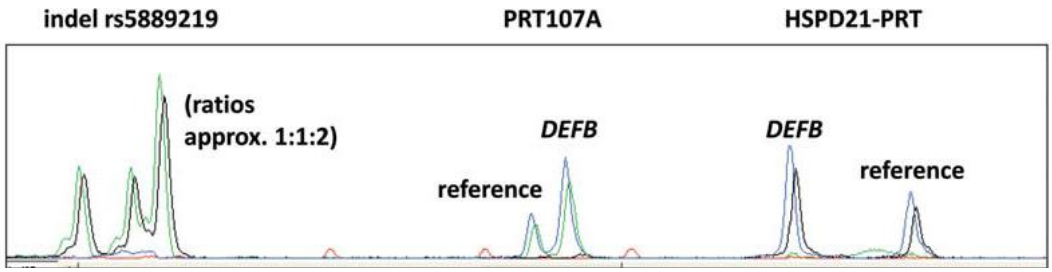


Figure 7. Indel and PRT amplification products.

In the electropherogram are represented indel and PRTs amplification products from the β-defensin multiplex PRT analysis. The approximate 1:1:2 ratios of alleles for the multi-allelic indel rs5889219 is consistent with a total of four copies, estimated by the ratio between the test (DEFB) and the reference product of the two PRTs (PRT107A and HSPD21-PRT). [Adapted from Aldhous et al. 2010].

Compared to other PCR based methods, PRT offers high accuracy in CNV calling, comparable to microarray analysis (Aldhous *et al*. 2010; Fode *et al*. 2011). PRT is suitable for CNV association studies as a rapid and inexpensive CNV genotyping method.

1.1.2.5 Future directions

The above techniques are significantly changing our point of view regarding the human genome, which cannot be defined by a single reference genome but can only be described by a dynamic series of variants, moreover, no single existing methodology can be used to analyse absolute copy number for all genomic regions (Cantsilieris, Baird, White 2012).

Despite the great advancement and augmentation of analytical tools of investigation, much progress is still to be made in the detection of structural variations. Firstly, the use of "gold standard" controls for discovery and genotyping CNV methods is necessary to increase the reproducibility and verify the reliability of intra- and inter-methodological results (Aldhous *et al*. 2010; Pinto *et al*. 2011). Accurate analysis of the human genome revealed bias towards the identification of deletion CNVs: quantitative differences between deletion copy numbers are greater than duplications and multiallelic loci. In fact, distinguishing between four and five copy number, with a relative increase of 0.25, is a considerable challenge using current standard methodologies and complex CNVs require even more thorough investigation due to unclear information from the reference genome and erroneous use of the investigative tools (Hollox 2010; Hollox 2012). Therefore, the detection of multiallelic CNV has to be improved in order to completely understand the impact of CNV on the human genome and disease. Ultimately, the optimization of combined NGS technologies to a routinely laboratory approach could significantly increase our knowledge of the human genome and its variants, such as CNV.

1.1.3 Molecular impact of CNV on gene expression

Associated to phenotypic variants in disease and healthy populations (Lupski and Stankiewicz 2006), CNV, as other genetic variation, can directly affect gene expression altering gene dosage. Indirect effects can occur through position effects or downstream pathways, disrupting the gene reading frame or perturbing gene regulatory networks (Dermitzakis and Stranger 2006; Hurles, Dermitzakis, Tyler-Smith 2008; Reymond *et al*. 2007).

Comprehensive transcriptome analyses and gene expression studies investigate the effects of CNV either from a population point of view or in a disease context (Aldred, Hollox, Armour 2005; Hollox, Armour, Barber 2003; Perry *et al*. 2007; Pollack *et al*. 2002; Stranger *et al*. 2007). Analyses of expression levels of 14,925 transcripts with SNPs and CNVs in lymphoblastoid cell lines of 210 unrelated HapMap individuals associated change in copy number to about one-fifth of the detected genetic variation (Stranger *et al*. 2007). Studies on mice and rat transcriptosomes indentified a less strong but still significant correlation between CNVs and gene expression (Guryev *et al*. 2008; Henrichsen *et al*. 2009).

Several gene expression studies showed that the functional effect of CNV can be described with a simple model where a duplication or a deletion of the gene copy are respectively correlated with an increase or decrease of its expression (Hollox, Armour, Barber 2003; Lupski *et al*. 1991a; Perry *et al*. 2007). According to the mechanism of CNV formation, the deletion/duplication event can cause gene interruption by disruption of the coding sequence causing a loss of function (Nathans *et al*. 1986). Rearrangements between different genes or the same gene or their regulatory regions can initiate gene fusion resulting in a gain-of-function mutation (Lifton *et al*. 1992); this mechanism is more likely to occur in cancers associated to chromosomal traslocation (Mitelman, Johansson, Mertens 2007). CNV involving regulatory regions within a specific gene can have impact on the gene itself or cause pleiotropic effects; modulation of genes in CNV proximities can occur by a position effect or perturbations of the transcription machinery followed by genomic landscape modifications (alterations of chromatin structure, positioning of chromatin, physical dissociation of the transcription unit) (Velagaleti *et al*. 2005). Moreover, CNV formation can cause deletion of one allele that may uncover another recessive allele or functional polymorphism (Kurotaki *et al*. 2005).

1.1.5 Phenotypic effects of CNV

CNVs have been shown to be widely accountable for human evolution and genetic diversity between individuals. The relevance of CNV on the human genome has been underlined by several association studies which highlighted how CNV can have an effect on susceptibility to cancer, complex trait disorders, and be the main cause of the onset of Mendelian diseases or confer a benign phenotype (Almal and Padh 2011; Stankiewicz and Lupski 2010; Zhang *et al*. 2009).

1.1.5.1 Mendelian disease

Frequently genetic rearrangements arisen after NAHR events result in genomic disorders such as Mendelian traits, contiguous gene syndromes and whole-arm chromosome aberrations (Stankiewicz and Lupski 2002; Zhang *et al*. 2009). Historically, the Charcot-Marie-Tooth disease type1A and hereditary neuropathy with liability to pressure palsies were the first disorders detected to be caused by CNV events, more specifically the copy number change of the dosage-sensitive gene *PMP22* as a result of the duplication of a

large region in chromosome 17p12 (Lupski *et al*. 1991b). Smith-Magenis syndrome / Potocki-Lupski syndrome is another genomic disorder with Mendelian autosomal dominant inheritance resulting from haploinsufficiency of the *RAI1* gene on chromosome 17p11.2 (Slager *et al*. 2003). Likewise, recessive dominant disorders or X-linked disorders are associated to CNVs. For instance, intrachromosomal rearrangements of the long arm of chromosome X causing recombination of the factor VIII gene are associated to haemophilia A pathology (Gaucher and Mazurier 1995).

1.1.5.2 Infectious and Autoimmune diseases

CNVs are enriched for genes with immunological functions, however the exact way in which they contribute to the genetic variability responsible for the recognition repertoire of the immune system is not explicitly understood (Olsson and Holmdahl 2012).

Fc γ receptors (*FcγR*) are low-affinity receptors for the Fc domain of immunoglobulin G, whose activation stimulates phagocytosis and release of cytokines and other inflammatory mediators, depending on the cell type. CNVs within this locus have been studied in association with several autoimmune diseases giving contradicting results. A recent meta-analysis of *FCGR3B*, encoding for a receptor variant expressed on neutrophils, mast cells, and eosinophils, confirmed the role of its low copy number with systemic lupus erythematosus (SLE) (McKinney and Merriman 2012).

Located in major histocompatibility complex (MHC) region, complement component C4 (*C4A* and *C4B*) encodes for pivotal component in the activation cascades of complement pathways, essential in the binding of the immune complex to the target (Carroll 1998; Mackay, Rosen, Walport 2001). Low and high *C4A* copy number are associated respectively to increased risk and protection to SLE, increased *C4B* copies with higher risk of rheumatoid arthritis (Lv *et al*. 2011; Rigby *et al*. 2012; Yang *et al*. 2007).

β-defensin locus, a copy number variable region strictly correlated with autoimmunity, will be discussed in 3.1.

1.1.5.3 Population genetics and evolution

CNV, as a structural change that significantly contributes to phenotypic variance, is likely to give a relevant signature to evolution and differentiation in modern humans

(Hurles, Dermitzakis, Tyler-Smith 2008). In particular, CNVs overlapping genes are more likely to be adaptive and affect fitness than in intergenic CNVs (Cooper, Nickerson, Eichler 2007). However, most CNVs located in functional regions in the human genome appear to be under negative selection (Nguyen *et al*. 2008), and only a few CNVs empirically showed positive selection in the population (Hardwick *et al*. 2011; Kidd *et al*. 2007; Perry *et al*. 2007; Xue *et al*. 2008).

The first CNV event to be studied showing relevant interest in evolutionary change is the salivary amylase gene (*AMY1*). Higher copy number of *AMY1*, correlated with different salivary amylase protein levels, is on average present in individuals from populations with high-starch diet than those with low-starch diet (Perry *et al*. 2007).

Another relevant gene showing CNV event undergoing adaptive evolution is the UDP-glucuronosyltransferase 2B17 (*UGT2B17*) gene. *UGT2B17* CNV associated with variation in urine testosterone level, fat mass, male insulin sensitivity and probably prostate cancer risk, differs in mean number between populations from Africa, Europe, and East Asia due to positive selection (Xue *et al*. 2008). The *UGT2B17* deletion, that was found to be highly common in Asian individuals, is probably responsible for conserving the testosterone level, generally lower in this population (Iskow, Gokcumen, Lee 2012).

A population genetics study has been done in regards of a complex CNV region, the β-defensin region, which encodes for important players in the host defense. From the analysis of CNV frequencies in 68 populations, within a worldwide distribution, it appeared that there are significant differences between populations and continents, considered to be the result of recent selective or demographic events (Hardwick *et al*. 2011).

1.1.6 CNV and neurodegenerative diseases

Genetic mutations located in unrelated genes have been linked to the familial versions of several of neurodegenerative disorders, including Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic lateral sclerosis (ALS). CNV as a cause of rearrangement of the genome structure could play an important role in the pathogenesis of these disorders.

AD is the most common neurodegenerative disorder of the elderly and is characterized by dementia and progressive cognitive decline. As a familial (1% of the cases)

or sporadic disorder, AD is attributed to rare mutations in several genes: the amyloid-β protein precursor (*APP*), presenilin 1 (*PS1*) and presenilin 2 (*PS2*), and the apolipoprotein E (*APOE*) ε4 allele, which are directly related to overproduction or reduced clearance of the amyloid-β peptides (Rogaeva, Kawarai, George-Hyslop 2006). Duplication events involving the *APP* gene have been shown to be associated with autosomal dominant early onset AD (ADEOAD) and also cerebral amyloid angiopathy (Rovelet-Lecrux *et al*. 2006). Five different types of duplication causing an accumulation of the APP peptide were identified causing ADEOAD. Interestingly, early onset AD is a common feature of Down syndrome due to the trisomy of chromosome 21, where the *APP* gene is situated (21q21.3). Several studies extensively investigated the impact of CNVs in AD and have shown numerous associations such as a duplication events on chromosome 15q11.2, copy number change of olfactory receptor genes cluster (14q11.2), CNV of complement receptor one (*CR1*) and, specifically in regard of sporadic ADEOAD cases, seven new rare CNVs (four of which involves genes (*KLK6*, *MEOX2*, *SLC30A3*, *FPR*) related to amyloid-β peptide metabolism or signalling) (Brouwers *et al*. 2011; Ghani *et al*. 2012; Heinzen *et al*. 2010; Rovelet-Lecrux *et al*. 2011; Shaw *et al*. 2011; Swaminathan *et al*. 2011).

PD is a progressive neuromuscular degenerative disease that affects 1% of the population over 50 years of age. The genetic etiology of PD is quite complex and involves several genes, at least five of which showed mutations responsible for autosomal dominant or autosomal recessive onset of the disease (Pankratz and Foroud 2007). Triplication and duplication of the *SNCA* gene, encoding for the α-synuclein, have been reported in individuals affected by familiar form of PD, suggesting an increase of α-synuclein due to CNV-induced dosage effect (Farrer *et al*. 2004; Fuchs *et al*., 2007; Ibanez *et al*. 2004; Miller *et al*. 2004; Singleton *et al*. 2003). A genome scan of CNV in a familiar PD cohort identified CNV in two novel genes (*DOCK5* and *USP32*) associated with an increased risk of PD (Pankratz *et al*. 2011). Exome screening in Brazilian patients with PD revealed parkin (*PARKIN*) and PTEN induced putative kinase 1 (*PINK1*) genes copy number variations, but no dosage alteration was found in *SNCA* and *DJ-1* genes (Moura *et al*. 2012). Furthermore, alterations in copy number of the microtubule-associated protein tau gene *(MAPT*) and of other genes located in its proximity (17q21.31) encode for a broad range of mental dysfunctions associated to variable phenotypes, ranging from frontotemporal dementia

with parkinsonism to mental retardation with or without dysmorphism and to schizophrenia (RoveletLecrux and Campion 2012).

ALS is a fatal neurodegenerative disease affecting the motor neurons resulting in progressive weakness and death, usually within about three to five years post onset, as a result of neuromuscular respiratory failure. Familial forms of ALS, which represent 10% of the total cases, have been associated with mutations in several genes such as superoxide dismutase-1 (*SOD1*), accounting for about 20 to 30% of onset (Al-Chalabi *et al*. 2012). Screening on large ALS sample cohort by MLPA analysis showed that CNV, i.e. duplications, of the survival motor neuron gene (*SMN1*), which homozygous deletion causes spinal muscular atrophy, represents a risk factor for sporadic ALS (Blauw *et al*. 2012).

1.2 Huntington's disease

"Chorea is essentially a disease of the nervous system": this is how George Huntington initiated his essay "On Chorea", in 1872. In this article, Huntington described accurately an already known disease, the chorea or dancing disorder; all the main symptoms and an illuminating description on the heritability are reported. This remarkable characterization of the disease led to the designation of the disease as Huntington's disease.

HD is a fatal progressive degenerative brain disorder, which is clinically described by motor dysfunction, severe emotional / psychiatric disturbances, cognitive decline and dementia (MacDonald *et al*. 1993). The manifestation of the disease symptoms is around the age of 40 years, but juvenile (onset <20 years) and older onset (>70 years) forms also exist. Although useful for diagnosis, movement impairment is a poor marker of disease severity and it may not be the main debilitating aspect of the disease (Walker 2007). HD carriers, in fact, show a wide range of phenotypic markers, including cognitive decline, personality change, anxiety, depression, irritability, sleeping disturbance, metabolic symptoms, all signs of high heterogeneity of the disease (Novak and Tabrizi 2011). After the diagnosis of a first visible symptom, HD mutation carriers have an average life expectancy of around 20 years.

HD manifests in all population groups with great variability between ethnic groups and geographical distribution: the incidence of HD in Asia varies from 0.046 to 0.16 per 100,000 per year, whereas incidence in Europe, North America, and Australia varies from 0.11 to 0.8 per 100,000 per year, within any trends toward a decrease or increase in incidence over time. The prevalence in the North America, Europe, and Australia is about 5.70 per 100 000 and the rate is much lower than 0.5-0.4 per 100 000 in Asia and Africa (Harper 1992; Pringsheim *et al*. 2012). The geographic difference in prevalence in HD can be explained by predisposing haplogroups model, albeit another model based on a founder

chromosome could be responsible for the variable expansion of the disease (Falush 2009; Squitieri *et al*. 1994; Warby *et al*. 2009).

1.2.1 Neuropathology of HD

HD is a neurodegenerative disease that affects all the brain but with prominent cell loss and atrophy within the striatal part of the basal ganglia, in the caudate and putamen (Vonsattel and DiFiglia 1998); this degeneration is followed by extensive astrogliosis, with an increasing severity in time, leading to a characteristic great enlargement of the lateral ventricles (Figure 8).

Recently the analysis of presymptomatic and early stage HD mutation carriers has permitted the evaluation of early effects of the disease on the brain, detectable through the application of brain imaging techniques including structural magnetic resonance imaging (MRI) or positron emission tomography (PET), able to highlight macro-structural changes such as atrophy, micro-structural alterations such as demyelination, metabolic dysfunctions and functional changes (Georgiou-Karistianis *et al*. 2013).

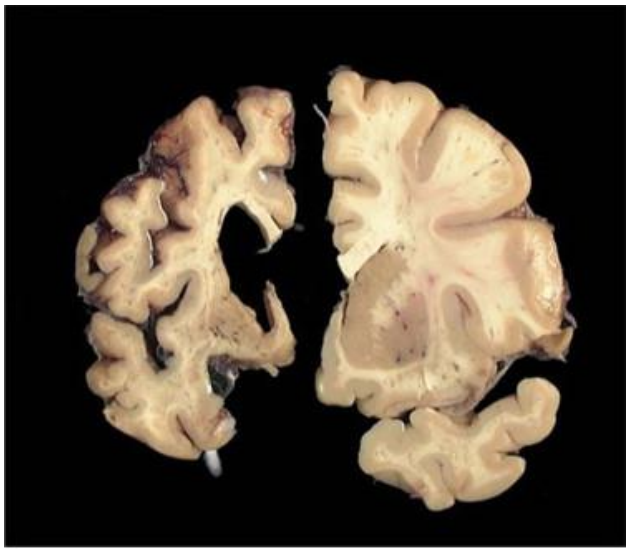


Figure 8. HD and normal brains.

On the left it is visible the great impact of HD within a profound shrinkage of cortex and caudate and the resulting ventricular expansion. Normal brain on the right. [From Reiner, Dragatsis, Dietrich 2011].

Large longitudinal studies (PREDICT-HD and TRACK-HD (Paulsen *et al*. 2006; Tabrizi *et al*. 2009) have shown that the reduction of the striatal volume is evident from up to 15 years prior to symptoms onset and that rates of decline in striatal volume is significantly faster compared with age-matched controls (Aylward *et al*. 2012; Paulsen *et al*. 2010; Tabrizi *et al*. 2011). Nevertheless other features including the whole brain atrophy, grey and white matter loss and cerebral spinal fluid (CSF) expansion have been detected as early signs of HD neurodegeneration, the striatal atrophy is the most valuable biomarker of HD progression (Georgiou-Karistianis *et al*. 2013).

1.2.2 The genetics of HD

HD is a monogenic disorder caused by autosomal dominant inheritance of an abnormal cytosine-adenine-guanine (CAG) trinucleotide expansion within the 5' end of the huntingtin (*HTT*) gene, located on the short arm of chromosome 4 at 4p16.3. The CAG triplet is the genetic code for the amino acid glutamine, which forms a toxic polyglutamine tail attached to the huntingtin protein in HD (MacDonald *et al*. 1993; The Huntington's Disease Collaborative Research Group 1993).

The disease shows incomplete penetrance when the mutation stretch is between 36 and 40 repeat units in *HTT* gene, a threshold beyond which the disease is considered fully penetrant (Figure 9) (McNeil *et al*., 1997; Quarrell *et al*. 2007; Rubinsztein *et al*. 1996). The length of this repeat expansion is significantly correlated with the age of onset (AO), with which it is inversely correlated (Figure 10) (Andrew *et al*. 1993; Brinkman *et al*. 1997; Duyao *et al*. 1993; Lucotte *et al*. 1995; Snell *et al*. 1993). However, in individuals with repeat units ranging between 36 and 50, the mutation length is a poor predictor for AO. In fact, most of the correlation is found with high repeat lengths > 60, which are associated with juvenile onset (Figure 9).

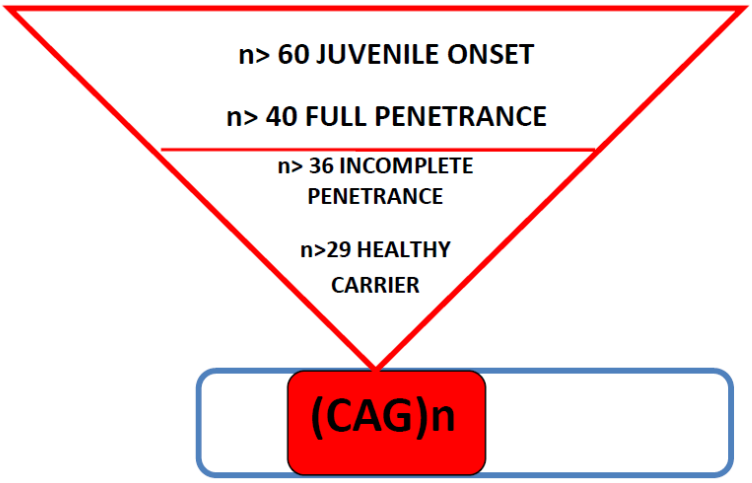


Figure 9. Mutation length in the *HTT* gene and disease onset risks.

HD, as other triplet disorders, shows the phenomenon of "genetic anticipation". Individuals homozygous for the *HTT* allele with less than 36 CAG repeat units are asymptomatic but, in case of the expansion size is between 29 and 36 units, there is a high risk for the progeny to manifest the disease. This is due to a pathogenic expansion of the CAG repeat caused by the inherent instability of DNA replication of repeat regions, especially in spermatogenesis. This effect is the cause of between 6-8 % of new cases of the sporadic form of HD. This effect is much more evident in the offspring of affected individuals: in this case there is anticipation of the disease in the offspring, who manifest the HD at earlier AO than the affected parent. In most of cases, subjects with juvenile onset of HD inherited the mutant allele from the paternal side (Kremer *et al*. 1995; Ranen *et al*. 1995; Trottier, Biancalana, Mandel 1994).

Instability of the CAG repeats occurs also in somatic cells in HD: evidence of mosaicism have been showed in selective area of the brain in humans, in particular correlated to long repeats stretch, and in mouse model of the disease (Kennedy *et al*. 2003; Telenius *et al*. 1994).

1.2.2.1 The age of onset in HD

AO in HD refers to the moment when a carrier of the mutated gene exhibits explicit HD signs. The accurate determination of AO is critical to identify factors that modify AO or biomarkers of disease progression, and to develop and evaluate therapies that aim to delay it. In manifesting cases, the AO is estimated retrospectively by clinicians based on information reported by the affected individual. In presymptomatic stages, mainly in familial cases, the expected age of future onset is calculated considering the CAG repeats length and the actual age of the individuals and verified prospectively. In both case prediction of the AO are susceptible to inaccuracies of the rater and variability due to the formula used for the estimates, based on correlation between AO and CAG repeat length (Langbehn *et al*. 2004; Langbehn, Hayden, Paulsen 2010; Orth and Schwenke 2011).

AO is estimated to range between 1 to 80 years. This wide temporal window can be partially explained by the inverse correlation with the variable length of the CAG expansion, which is best described by a curvilinear regression model, using a log-transform of AO (Figure 10). From this correlation arises that the CAG mutation length accounts for nearly 70 % of the variance of AO, albeit in individuals carrying 40-50 CAG repeats the rate decreases significantly to 44% suggesting the presence of other modulators (Wexler *et al*. 2004). These could be stochastic, environmental / experiential or genetic factors. A study performed on 83 independent Venezuelan HD pedigrees estimated the impact of additional genetic, and environmental, factors on the AO in HD. After excluding any possible influence of shorter *HTT* allele or gender, it was estimated that the residual AO variance in HD was 63 % attributable to shared (family) and unshared environments, and the remaining 38 % to genetic factors (Wexler *et al*. 2004). Non-influential is the presence of a second mutant *HTT* allele as homozygotes for HD show comparable AO to heterozygotes, although the disease progression is much more severe (Myers *et al*. 1989; Squitieri *et al*. 2003; Wexler *et al*. 1987).

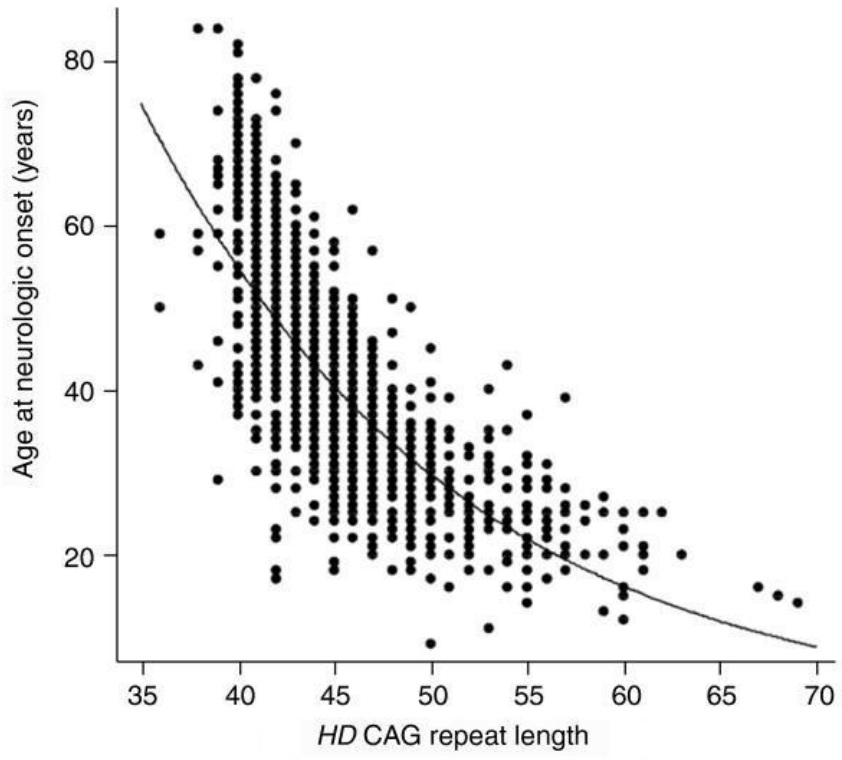


Figure 10. The inverse relationship between age of onset and repeat length.

Box plot of AO and CAG repeat length of the longer allele. The curvilinear relationship between the two variables can be observed. It also is important to note the wide variability of AO within each repeat length, especially within the range from 40 to 50 CAG repeat units. [From Gusella and MacDonald 2009].

A genetic modifier of the AO in HD is a gene harbouring a mutation which both influence the onset of phenotypes associated with the CAG mutation and affect gene function. The main strategy to identify genetic modifiers in HD is usually the candidate approach: common genetic variants, likely involved in HD pathogenesis, are analysed in manifest HD carrier cohorts to test whether they are associated with earlier or later AO among similar CAG length carriers. So far, several screens on HD sample cohorts have been performed, giving a mixture of results. The table 2 lists some of the main loci or gene variants investigated to be genetic modifiers of AO in HD. Several recent studies performed, aimed to validate previous findings in large cohorts, do not replicate the association of the listed variants with the AO in HD, initially analysed on small sample sets.

Table 2. Association studies of genetic modifiers of the AO in HD.

Gene ID	Analysed Variant	Study	Population	analysed HD samples*	protective / detrimental allele	p value
GRIK2	(TAA)n	Rubinsztein *et al*. 1997	England	293	all others / (TAA)16, 17	0.008 (*ΔR^2 =0.41)*
		MacDonald *et al*.1999	USA	258		0.009 (*ΔR^2 =0.006)*
		Chattopadhyay *et al*. 2003	India	77		0.009 (*ΔR^2= 0.02)*
		Cannella *et al*. 2003	Italy	524		0.001* (*ΔR^2= 0.14)*
		Metzger *et al*. 2006a	Europe	980		ns
		Naze *et al*. 2002	France	138		ns
		Andresen *et al*. 2007	Venezuelan kindreds	368		ns
		Lee *et al*. 2012a	HD-MAPS, REGISTRY-EHDN, COHORT	2,911		ns
GRIN2A	rs2650427	Arning *et al*. 2007	Germany	250	T/C-additive model	0.047*(*ΔR^2=0.011)*
		Andresen *et al*. 2007	Venezuelan kindreds	421		0.04 (*ΔR^2=0.003)*
		Saft *et al*. 2011	REGISTRY-EHDN	1,211		0.028 (*ΔR^2= 0.001)*
GRIN2B	rs1806201	Arning *et al*. 2007	Germany	250	T / C	0.005* (*ΔR^2=0.026)*
		Andresen *et al*. 2007	Venezuelan kindreds	443		ns
		Saft *et al*. 2011	REGISTRY-EHDN	1,211	CT / all others	0.046*(*ΔR^2=0.003)*
PGC-1α	rs6821591	Weydt *et al*. 2009	Italy	447	G/A dominant model	0.0178
	rs7665116				G/A additive, dominant model	0.0016, 0.0003
		Taherzadeh-Fard *et al*. 2009	Germany	400	C/T dominant model	0.012 (*ΔR^2=0.004)*
		Che *et al*. 2011	Europe	854	additive, dominant model	0.03 (*ΔR^2=0.0129)*, 0.0112 (*ΔR^2=0.0128)* (0.0172; 0.0045)*
		Ramos *et al*. 2012	Southern European, Western European	1,628	C/T additive, dominant model	0.009(*β=0.090),* 0.008(*β=0.113).* [0.047(*β=0.069)*, 0.033(*β=0.089)*]**

UCH-L1	Ser18Tyr	Naze *et al*. 2002	France	138	Y/S dominat model	0.024 (*ΔR^2=0.042)*
		Metzger *et al*. 2006b	Europe	946	Y/ S	< 0.001*
		Andresen *et al*. 2007	Venezuelan kindreds	405		ns
BDNF	Val66Met (rs6265)	Alberch *et al*. 2005	Spain	122	VM heterozygous / VV homozygous	< 0.001 (*ΔR^2=0.59)*
		Metzger *et al*. 2006a	Europe	980		ns
		Mai *et al*. 2006	Germany	250		ns
		Di Maria *et al*. 2006	Italy	244		ns
		Kishikawa *et al*. 2006	HD-MAPS	557		ns
		Taherzadeh-Fard *et al*. 2010	Germany	419		ns
HAP-1	T441 M (rs4523977)	Metzger *et al*. 2008	European	903	MM / all others	0.015*
		Taherzadeh-Fard *et al*. 2010	Germany	419		0.041 *(*ΔR^2=0.002)*
		Karadima *et al*. 2012	Greece	298		ns
ADORA2A	rs5751876	Dhaenens *et al*. 2009	France	791	all others / TT	0.019
	rs5751876	Taherzadeh-Fard *et al*. 2010	Germany	419	TT ±	0.032*(*ΔR^2=0.003)*
	rs2298383				TT ±	0.006(*ΔR^2= 0.004).* 0.001*(*ΔR^2=0.007)*

** The statistic value refers to a specific subset.*

***Analysis adjusted for ancestry.*

± The genotype increased the R^2 of the regression CAG repeats-AO, with no variance in AO.

ΔR^2 indicates the improvement of the regression model when the various genotypes are considered in addition to the HD CAG repeats.

β value refers to the size of effect of the integration of the genetic variant to the model of correlation between CAG repeats-AO.

An example of a false-positive association due to a small sample size effect is the study of a variant on *GRIK2* (glutamate receptor, ionotropic kainate 2), gene encoding the GluR6 subunit of the main excitatory neurotransmitter receptor family in the brain, a valuable candidate possibly involved in excitoxicity in HD. A triplet expansion of 16 TAA

repeats in this gene was reported to be associated with earlier onset of HD in several studies, however in each study less than 300 individuals were analysed or the correlation was just present in a small subset (Cannella *et al*. 2003; Chattopadhyay *et al*. 2003; MacDonald *et al*. 1999; Rubinsztein *et al*. 1997). Further analyses clearly disproved this association analyzing the questioned variant on larger data sets, highlighting the need of large-scale investigations to adequately evaluate the potential effects of candidates genetic modifier of the AO of HD (Lee *et al*. 2012a; Metzger *et al*. 2006).

Another interesting example of a biased association is the case of *PGC1-α* peroxisome proliferator-activated receptor gamma, coactivator 1 alpha). Association studies of SNPs and haplotypes on this gene with a delay of AO in HD and much evidence of its active role during the neurodegenerative process supported the role of *PGC1-α* as a genetic modifier in HD (Chaturvedi *et al.* 2009; Taherzadeh-Fard *et al*. 2009; Weydt *et al*. 2006; Weydt *et al*. 2009). A further genetic screening on several HD cohorts highlighted that the strong effect seen in the previous studies was probably a result of population-dependent phenotype stratification. In fact, Southern European HD individuals, representing the majority of the previously analyzed samples, are enriched for the minor allele frequency (MAF) of one of the main SNP previously associated with delay of AO, and they are much more significantly associated to a later HD AO compared to Western European HD individuals. In addition to question the role of *PGC1-α* as a genetic modifier of the AO in HD, the results of this study reveals genetic ancestry as a critical factor in HD association studies (Ramos *et al*. 2012).

Considering these and other recent works, the analysis of genetic modifiers of AO in HD should strive towards standardized approaches. In order to avoid batch effect or stratification due to population dependent phenotypes it is fundamental to perform association studies on a large HD individual data set provided by multi-centre, multi-national observational studies like the “REGISTRY” project of the European HD network (EHDN), and the American “COHORT” study, which catalogues natural history data on a wide range of the HD population (Dorsey 2012; Orth 2011). In addition to that, it is important to screen for a statistically well-behaved data set that conforms to the primary assumptions of linear regression analysis (constant variance and normally distributed error) between AO and

triplet mutation length. In fact, a single outlier can potentially have a huge effect on the inferences between the two variables, AO and CAG repeat length, misleading the evaluation of a potential modifier of the AO (Lee *et al*. 2012b). Moreover, a genetic association study should provide evidence of how the studied variant affects gene function and HD onset.

Currently, genetic variants, which can be reliably considered modifiers of the AO in HD, are few and independent replication studies are still needed. However, noteworthy is the association study in regard of *HAP-1* (huntingtin associated protein 1): a polymorphic variant in *HAP-1* was found to be responsible for a change in the protein structure leading to a delay of motor symptoms onset in individuals with < 60 CAG. This finding was validated by *in vitro* study where the polymorphic protein was associated with a reduction of HTT-toxicity (Metzger *et al*. 2008). Two further studies weakly reproduced and failed to reproduce respectively this previously observed association; they were however performed on smaller cohorts (Taherzadeh-Fard *et al.* 2010; Karadima *et al.* 2012).

Confirming the Wexler *et al.* study, it has been found that the shorter / wild type allele is not influential on the onset and progression of motor symptoms in HD, a point of debate in several studies (Lee *et al*. 2012b), and any possible implication of *HTT* region SNP haplotypes with altered CAG repeat length distribution or residual AO of HD has been excluded (Lee *et al*. 2012c).

An alternative approach for the discovery of possible genetic modifiers of AO in HD makes use of screen in simple and manageable model organisms as yeast, *Caenorhabditis elegans* and *Drosophila melanogaster* in which human mutant *HTT* is introduced. Such models are useful for the identification of modifiers of the effects of this fragment (Giorgini *et al*. 2005; Kaltenbach *et al*. 2007; Lejeune *et al*. 2012). Limitations of this approach lie in the diversity of HD expression in these organisms and in the necessary replication of possible findings in other models and human genetic screening. In order to reach a deeper knowledge and holistic view of all genomic variants (not only SNPs) associated to HD, a significant advantage will be provided by genome wide and exome wide sequencing and array studies with high quality annotation, performed on thoroughly understood cohorts.

This will provide a better platform from which to further interrogate potential modifiers of AO in HD.

1.2.3 Huntingtin protein

The gene encoding for the huntingtin protein (HTT) lies on a region of ~170 kb at 4p16.3, consisting of 67 exons. It is quite conserved among vertebrates, where the most divergent species shares 69% identity at the nucleotide level (Sathasivam *et al*. 1997). Huntingtin is a completely soluble protein of 3,144 amino acids, whose structure is still under investigations due to the large size and difficulties in creating crystals. The gene encodes for two alternatively polyadenylated transcripts, variable in size (~3kb) and localization: the shorter form is ubiquitously expressed, whereas the longer HTT is predominantly present in the neurons of the CNS, where the expression of the gene is enriched (Li *et al*. 1993; Lin *et al*. 1993; Strong *et al*. 1993). Specifically, HTT is highly localised in cortical pyramidal neurons in layers III and V that project to the striatal neurons (Fusco *et al*. 1999).

HTT has been shown to be expressed from the embryonic stage, during which mutated forms or reduced expression are tolerated but not the complete deletion, which causes early embryonic lethality in mice (Duyao *et al*. 1995; Myers *et al*. 1989; Nasir *et al*. 1995; Wexler *et al*. 1987; Zeitlin *et al*. 1995). At cellular level, the HTT protein is associated with a range of organelles including plasma membrane, endocytic and autophagic vesicles, endoplasmic reticulum, endosome, Golgi, and mitochondria (DiFiglia *et al*. 1995). Within neurons, it is present in neurites and associated with vescicular compartments at the synapses. This extensive subcellular localization does not facilitate the characterization of the HTT protein function (Zuccato, Valenza, Cattaneo 2010).

Organized into three clusters, 16 HEAT (Huntingtin, Elongator factor3, the regulatory A subunit of protein phosphatase 2A, and TOR1) domains are present in the HTT protein (Figure 11) (Tartari *et al*. 2008). HEAT repeats consist of ~50 amino acids which are conserved in several proteins and are organized in tacked pairs of antiparallel alpha-helices; they are involved in protein–protein interactions (Andrade and Bork 1995).

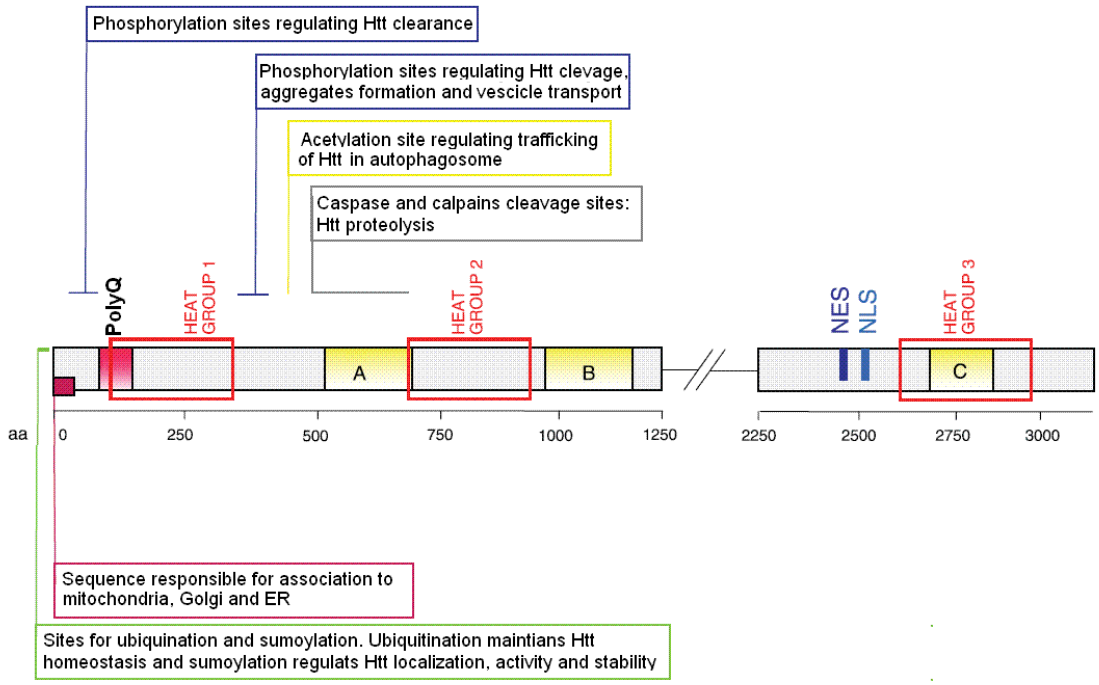


Figure 11. Schematic diagram of the Htt protein sequence.

The pink box is a regulator of HTT binding to other proteins. PolyQ indicates the polyglutamine tract. The red empties boxes indicate the three main groups of HEAT repeats. The yellow boxes are tissue-specific cleavage sites of HTT: B, regions cleaved mainly in the cerebral cortex; C, regions cleaved in the striatum; A, regions cleaved in both. NES is the nuclear export signal and NLS is the nuclear localization signal. aa amino acid, HTT huntingtin, ER endoplasmatic reticulum. [Adapted from Zuccato, Valenza, Cattaneo 2010].

The HTT protein exhibits consensus cleavage sites for the proteolytic cut of several proteases including caspases, calpain and matrix metalloproteinases, which likely generate a wide range of NH_2-terminal fragments (Figure 11) (Gafni and Ellerby 2002; Miller *et al*. 2010; Wellington *et al*. 1998). Despite that the wild type HTT is found in the cytoplasm, these NH_2-terminal fragments and the mutant HTT tend to shift and accumulate in the nucleus, a toxic event in HD Davies *et al*. 1997; Hermel *et al*. 2004; Landles *et al*. 2010). The inhibition of some of these enzymes is associated with a reduction of the toxicity, underling the importance of proteolysis of the mutant HTT full-length in HD Gafni *et al*. 2004; Wellington *et al*. 2000). Located at the COOH-terminal HTT site, the nuclear export signal mediates the nucleus-cytoplasm shuttle of HTT. This domain is cleaved away in mutant HTT, suggesting it

contributes to its accumulation in the nucleus, (Xia *et al*. 2003). A nuclear localization site at the NH_2-terminal HTT fragments binds the nuclear pore protein, involved in nuclear export, which is disrupted in HD due to the polyQ expansion interference (Cornett *et al*. 2005).

Moreover, HTT is subjected to posttranslational modifications, including ubiquitination, phosphorylation, acetylation, sumoylation, and palmitoylation. These could regulate its stability, localization, and function and be influential during HD pathogenesis, reducing or exacerbating the mutant HTT toxicity (Aiken *et al*. 2009; DiFiglia *et al*. 1997; Kalchman *et al*. 1996; Steffan *et al*. 2004).

1.2.3.1 Htt functions

HTT plays a key role during embryogenesis, as its complete inactivation in HTT knock-out mice causes embryonic death (Duyao *et al*. 1995; Nasir *et al*. 1995; Zeitlin *et al*. 1995). The absence of HTT leads to an increase in cellular death, as the anti-apoptotic function of the wild-type protein is lost (Zheng and Diamond 2012). In fact, it was found that the overexpression of wild-type HTT in brain-derived cells has a pro-survival function, whereas a depletion leads to higher response to apoptotic signals in neurons (Leavitt *et al*. 2006; Rigamonti *et al*. 2000; Zhang *et al*. 2003).

HTT is implicated in several transcription pathways. It is considered to be a primary regulator in the production and trafficking of the brain-derived neurotrophic factor (BDNF) in the cerebral cortex (Gauthier *et al*. 2004; Zuccato *et al*. 2001; Zuccato and Cattaneo 2009). Specifically, HTT regulates the activity of one of the BDNF suppressors, the repressor element-1 silencing transcription factor (REST) /neuro-restrictive silencer factor (NRSF) with which HTT interacts in a complex with other proteins, as HAP-1 (Zuccato *et al*. 2003; Zuccato *et al*. 2007).

Moreover, HTT as a membrane-scaffolding protein interacts with a wide range of endocytic / trafficking proteins and plays an important role in axonal vesicles transport and synaptic plasticity (Caviston *et al*. 2011; Trushina *et al*. 2004; Velier *et al*. 1998).

A better understanding of HTT functions comes from the analysis of mechanisms behind HD (1.2.4), where the mutant protein is a key player.

1.2.4 Mechanisms of neurodegeneration in HD

There are several mechanisms that contribute to the neurodegeneration observed in HD which are not exclusively correlated to HTT altered functionality. This suggests that HD is probably a non cell autonomous disease where the mutated HTT is the triggering factor in the disease.

1.2.4.1 Misfolding and Aggregation of the Mutant Huntingtin

At a cellular level, one of the first hallmarks of HD is the accumulation of N-terminal fragments. These fragments are mainly formed after the cleavage by several caspase, calpain and other proteases enzymes on both normal and mutant HTT, although the mutant form is more prone to the proteolytic cut and N-terminal fragments can produced by aberrant splicing CAG length-dependent of the *HTT* transcript (Kim *et al*. 2001; Lunkes *et al*. 2002; Sathasivam *et al*.„ 2013). From the cleavage of the mutant HTT, NH_2-terminal fragments accumulate in the nucleus of non neuronal and neuronal cells, where they can aggregate in nuclear inclusions (DiFiglia *et al*. 1997; Goldberg *et al*. 1996). Interestingly, it has been shown that the inhibition of the proteolytic enzymes or mutation in the cleavage consensus site on the *HTT* gene slow down the progression of HD in mice (Gafni *et al*. 2004; Graham *et al*. 2006; Wellington *et al*. 2000).

HTT fragments tend to aggregate into ultra-structures and deposit in neurons of all cortical layers and in the striatum of HD patients. Critical factors in the formation of the aggregates are the length and amount of the fragments, the length of the polyglutamine tail attached to them, and the presence of intracellular protein interactors of HTT Chen, Ferrone, Wetzel 2002; Hackam *et al*. 1998; Li and Li 1998; Martindale *et al*. 1998). These aggregates occur in different structures: oligomers, fibrils or amorphic aggregates, which can all form inclusion bodies (Figure 12).

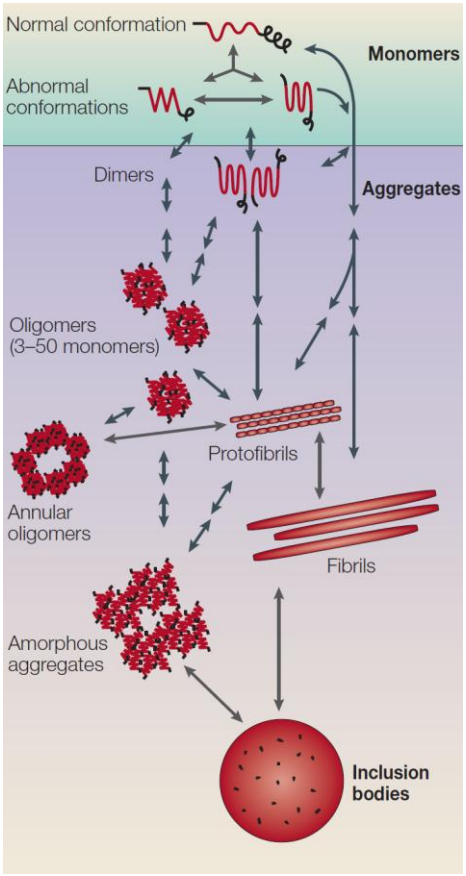


Figure 12. Possible aggregation pathway and formation of inclusion bodies.

[From Ross and Poirier 2005].

Although intranuclear deposits of mutant HTT alter gene expression in HD animal and cell models, the specific role that the nuclear inclusions play in HD pathology is controversial. These particles are known to be toxic with the potential to cause axonal degeneration or to recruit a wide range of factors, causing disruption of axon physiological activity, but can also be neuroprotective for the ability to sequester toxic soluble fragments or other elements (Zuccato, Valenza, Cattaneo 2010). Another hypothesis defines the small soluble monomeric forms alone toxic but not oligomeric or aggregated mutant HTT (Lu and Palacino 2013). It is possible to say clearly that the toxicity of these aggregates is due to their formation process more than to a single species *per se*; the misfolding process at the beginning is the evident cause of toxicity generated from the mutant HTT. A better understanding of the role of aggregation will be achieved within the study of early stage

formation events and, possibly, performing *in vivo* experiments (Ross and Poirier 2005; Zheng and Diamond 2012).

1.2.4.2 Clearance of Mutant HTT

The two main mechanisms involved in the clearance of intracellular polyQ protein deposits are the ubiquitin-proteasome system (UPS) and the autophagy.

HD being a proteinopathy, several studies have investigated the role of the UPS in the disease, showing that HTT protein aggregates are positively stained for ubiquitin in human HD post-mortem brains, HD cells and animal models (Schipper-Krom, Juenemann, Reits 2012). According to an early study, HTT aggregates can directly inhibit the UPS, which is impaired in HD (Bence, Sampat, Kopito 2001; Holmberg *et al*. 2004). Conversely, it was shown that the UPS inability to completely process long polyQ proteins might lead to the formation of aggregation-prone polyQ peptides, suggesting an active role of the UPS in the pathogenesis of HD, or the inactivation of the proteasome itself, due to clogging by the undigested polyQ (Holmberg *et al*. 2004; Raspe *et al*. 2009; Venkatraman *et al*. 2004). In another model the nuclear inclusions in striatal neurons restore the UPS and they probably interact with and thereby sequester toxic forms of mutant HTT with beneficial effects (Arrasate *et al*. 2004; Mitra, Tsvetkov, Finkbeiner 2009). According to these studies, it is not possible to define the role of the UPS but results clearly show that enhancement of proteasome degradation could stimulate and improve clearance of mutant HTT and prevent aggregation and toxicity of the polyQ fragments (Schipper-Krom, Juenemann, Reits 2012).

Initially observed in post-mortem brains (Tellez-Nagel, Johnson, Terry 1974), alterations in autophagy are typical in HD. The accumulation of aggregates in HD is shown to be responsible for the sequestration of mTOR, a negative regulator of the autophagic pathway (Ravikumar *et al*. 2004). The use of rapamycin, an inhibitor of mTOR, or the expression of small-molecule enhancer of rapamycin (SMER), enhancer of autophagy in yeast, attenuates the polyglutamine toxicity in various model of HD (Ravikumar *et al*. 2004; Sarkar *et al*. 2007). Inversely, soluble mutant HTT levels, aggregate formation, and toxicity augment when the autophagy is inhibited (Ravikumar, Duden, Rubinsztein 2002). Interestingly, HTT shares several similarities with mTOR: they are scaffolding proteins, both

of them containing HEAT repeats domains, important to specifically recognize membranous structures, and they both reversely associate to ER and Golgi membrane. Therefore, HTT is considered to have a potential role in the regulation of autophagy (Son *et al*. 2012).

1.2.4.3 Transcriptional Dysregulation

Initially, altered expression profiles of mRNA and protein levels of several neuronal peptides, evidenced by studies on *post mortem* HD brain and mouse HD model, supported the line that the transcriptional dysregulation is a causative factor of the neurodegenerative process in HD Andrews *et al*. 1999; Augood *et al*. 1996; Augood, Faull, Emson 2004; Cha *et al*. 1998; Cha *et al*. 1999). Further analysis of HD models (yeast, rat and mice) and HD human specimens by single gene and transcriptional profiling approaches contributed to identify a wide set of alterations in the expression of coding and non-coding RNAs (Borovecki *et al*. 2005; Ferrante *et al*. 2003; Giorgini *et al*. 2008; Hodges *et al*. 2006; Kita *et al*. 2002; Luthi-Carter *et al*. 2000; Luthi-Carter *et al*. 2002; Seredenina and Luthi-Carter 2012). These changes, mainly downregulation of gene expression, are mostly the result of the interaction between mutant HTT with several transcription factors such as CREB binding protein (CBP), TATA-binding protein (TBP), specificity protein 1 (Sp1), nuclear co-repressor (NCoR), and REST/NRSF and also p53, a tumour suppressor. Furthermore, the inhibition of enzymes involved in chromatin remodelling, and the interaction of HTT with components of the core transcriptional machinery and the RNA polymerase itself suggest the impairment of the transcriptional machinery in HD (Cha 2007).

Although most of the functional downstream effects derived from transcriptional dysregulation in HD represent an epiphenomena or have no relevant impact, some of them actively contribute to the disease pathology, causing for instance neuronal vulnerability (due to reduced expression of BDNF, NMDA and several other receptors and neuronal peptides) or metabolic dysfunctions (due to decrease of PGC1-α, a transcriptional coactivator).

1.2.4.4 Trafficking alterations

The ability of HTT to recruit (and sequester within the aggregates) several factors implies that intracellular transport might be impaired in HD. Several studies on mice and

Drosophila HD model shown that axonal transport disruption is associated with mutant HTT and occurs from the early stages of the disease, causing vesicle accumulation, typical deficit of axonal transport in HD (Gunawardena *et al*. 2003; Lee, Yoshihara, Littleton 2004; Li *et al*. 2001; Trushina *et al*. 2004).

HAP-1, which interacts with motor proteins as kinesin light chain and dynactin, is an essential interacting factor of HTT in trafficking (Engelender *et al*. 1997; Li *et al*. 1995; McGuire *et al*. 2006). In HD, mutant HTT-Hap-1 complex is mainly responsible for the disruption of the excitatory / inhibitory homeostasis, by altering the transport of NMDA and GABA receptors, and depletion of BDNF trafficking Fan and Raymond 2007; Gauthier *et al*. 2004; Twelvetrees *et al*. 2010).

Furthermore, alterations of the endosomal recycling are found in several models of HD due to dysfunction in the Rab protein family, with implications on the glucose uptake in primary neurons in the disease (Li *et al*. 2009a; Li *et al*. 2009b; Li *et al*. 2012). Interestingly, in a *Drosophila* HD model early synaptic impairments can be rescued by the overexpression of a member of this family, Rab11, potentially a good therapeutic target Steinert *et al*. 2012).

1.2.4.5 Metabolic dysfunctions

Metabolic abnormalities, specifically dysfunction in glucose metabolism, cholesterol biosynthesis and urea cyclic metabolism, are typical features present in HD human and mouse models (Chiang *et al*. 2007; Hurlbert *et al*. 1999; Josefsen *et al*. 2007; Shin *et al*. 2013; Valenza *et al*. 2005). The application of ^{1}H-magnetic resonance spectroscopy (^{1}H-MRS) and PET showed bioenergetic disfunctions in HD patients such as higher concentration of lactate in the cerebral cortex and basal ganglia (sign of an elevated glycolytic rate), or dysfunction of the glucose metabolism, present in presymptomatic and early stages of the disease (Ciarmiello *et al*. 2012; Feigin *et al*. 2001; Jenkins *et al*. 1993; Kuhl *et al*. 2004). Interestingly, the reduction of glucose uptake in the caudate was found to be an accurate biomarker of disease progression in presymptomatic individuals (Ciarmiello *et al*. 2012).

At cellular level disturbances of the energetic balance are mainly triggered by mutant HTT on the mitochondria (Panov *et al*. 2002). HTT interferes with calcium

homeostasis, production of ATP and also impairs mitochondrial trafficking, by direct effect on the mitochondria or modulating the expression of mitochondrial factors as PGC1-α, brain-type creatine kinase (CKB) or AMP-activated protein kinase (AMPK) (Ju, Lin, Chern 2012).

1.2.4.6 Neuroinflammation

Since high levels of activated microglia near degenerating neurons were detected in *post mortem* examination of HD brains, several studies investigated the role of neuroinflammation in HD, not only an epiphenomenom but probably a triggering element of the neurodegeneration (Meßmer and Reynolds 1998a; Meßmer and Reynolds 1998b; Singhrao *et al*. 1999). Further studies based on PET assessment estimated that activated microglia are manifest as early as 15 years before the predicted age of onset (Tai *et al*. 2007). Proteomic profiling of HD mutation carriers revealed the significant expression of components of the complement cascade and other cytokines in plasma and striatum of presymptomatic and early stage of the disease, indicative of the critical role of the innate immune system in HD (Björkqvist *et al*. 2008; Dalrymple *et al*. 2007).

The kynurenine pathway, initially known for one of its metabolites (quinolinic acid), which is able to induce HD like-symptoms in mice, via excitoxicity and free radical generation, played an important role neuroinflammation and also metabolic perturbation since early stage of HD (Tan *et al*. 2012; Amaral *et al*. 2013). Interestingly, the inhibition of a key enzyme (kynurenine 3-monooxygenase (KMO)) in the kyneurine pathway ameliorates neurodegeneration in a fruit fly and mouse models of HD (Zwilling *et al*. 2011; Thevandavakkam *et al*. 2010), findings that started the analysis of KMO inibhithors in clinical trials.

1.2.5 Models of HD

Initially HD animal models were produced by injecting neurotoxins (glutamate analogs, quinolinic acid, 3-nitropropionic acid and malonic acid) into the striatum, capable of reproducing various aspects of the disease (Zuccato, Valenza, Cattaneo 2010). After the

cloning of the human HTT, HD models were designed to mimic and study the effect of the polyglutamine expansion *in vivo* and *in vitro* (Table 3).

In vitro studies mainly contributed to explain the effect of aggregation of mutant HTT using different inducible or stable transformed cell lines such as the non-neuronal human HeLa cells, the human embryonic (HEK293T) and monkey kidney fibroblast cell lines (COS-7) as well as the Neuro2a (N2a) neuroblastoma (mouse) and neuron-like PC12 (rat) cells. Considering that mutant HTT is expressesed in several peripheral tissues, also immortalized cell lines from HD individuals specimens are used a reliable cell model for HD (Cisbani and Cicchetti 2012). A new promising approach is based on reverting skin fibroblast into inducible human pluripotent stem in order to direct these into neuronal cell development: with this method is possible to have direct specimens from HD individuals and to study the disease in untransformed human neuronal cells (Kaye and Finkbeiner 2013).

Table 3. Comparison of HD models available.

	Transformed HD cell lines	Human HD cell lines	Yeast	Nematode	Fruit fly	Zebrafish	Rodent
Availability	++	-/+	++++	++	++++	++	+
Cost	+++	+++	+	+	++	++	++++
Labour	++	+++	-	-	-/+	++	++++
Throughput	++	++	++++	++++	++++	++++	+
Homology	+++	++++	+	-/+	-/+	+	++
Pleiotropic studies	-	-	-	+	+++	+++	++++

In vivo approaches include the use of simple organisms and more sophisticated mammals. *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), and *Danio rerio* (zebrafish) and are all simple organisms suitable for economical and valuable discovery of new disease factors, validation of genetic screening and testing new therapeutic compounds. Transgenic and knock-in rodent lines,

especially *Mus musclulus* (mouse), are mammalian models that offer the study of neuronal systems closer to humans and allow the performance of complex behavioural test not possible in other models, but require more time and expense. Importantly, rodent models represent a key step in drug development pipelines. A subset of these models is discussed in detail below.

1.2.5.1 Yeast HD model

The budding yeast *S. cerevisiae*, thanks to its ease of manipulation and amenability to genetic modifications, large scale screening approaches, and a fully annotated genome, is widely used as a model organism in the study of neurodegenerative disorders (Pereira *et al*. 2012). The yeast does not contain an HTT ortholog however the expression of a human htt fragment with a polyQ tract, in the disease range size (HTT103Q), causes typical features of HD such as protein aggregation, mitochondrial dysfunction, defects in trafficking and transcription, resulting in apoptotic death and toxicity. All of these features make yeast a good disease model for HD, despite a main drawback: as unicellular organism the study of complex mechanisms as inflammation or synaptic transmission is not possible (Mason and Giorgini 2011). Interestingly, yeast HD model studies highlighted a link between HD and yeast prions, wherein the latter are necessary to trigger the polyQ toxicity (Duennwald *et al*. 2006; Meriin *et al*. 2002). A number of promising therapeutic compounds such as inhibitors of KMO (see 2.4.6) and the potent aggregation inhibitor C2-5 were identified in yeast screening and consequently validated in other HD models (Campesan *et al*. 2011; Chopra *et al*. 2007; Giorgini *et al*. 2005; Zhang *et al*. 2005; Zwilling *et al*. 2011).

1.2.5.2 Fruit fly HD model

The fruit fly is one of the most widely used genetic model organisms. Comparative genome wide sequence analysis of the whole fly genome, which contains only 4 pairs of homologous chromosomes and ~ 12,000 annotated gene, showed high structural similarity between fruit fly genes and human genes (Adoutte *et al*. 2000). As well a number of signalling and regulatory pathways and fundamental cellular processes (including gene expression, subcellular trafficking, synaptic transmission, synaptogenesis, and cell death) are well conserved or similar between human and fly (Ambegaokar, Roy, Jackson 2010).

The fruit fly is equipped with a specialized nervous system able to perform complex activities such as vision, olfactory processing, responsive motor behaviour (such as walking or climbing), learning and memory, similar to the mammalian brain (Marin *et al*. 2002; Rein *et al*. 2002; Wong, Wang, Axel 2002). Furthermore, the fly eye, characterized by a repetitive modular pattern (ommatidia) of photoreceptors (rhabdomeres), is a direct window of the neuronal integrity: i.g. neurodegenerative alteration can be easily recognizable by visible disruption of the ommatidial module under light microscope (Figure 13) (Jackson *et al*. 1998).

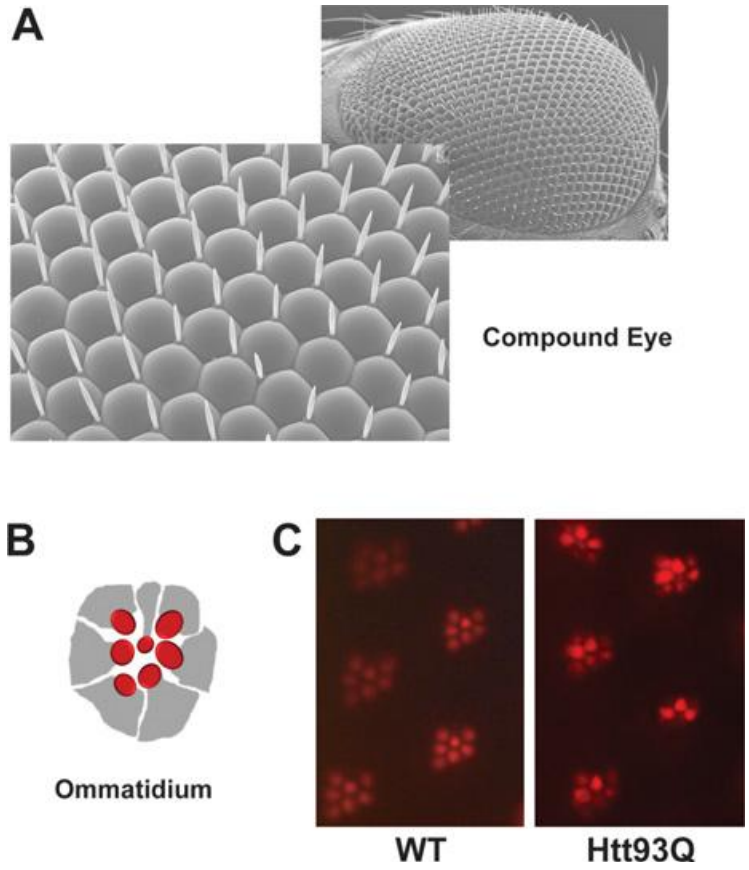


Figure 13. Ommatidia pattern and pseudopupil assay.

A)The repeating ommatidial pattern shown in scanning electron microscopy images of the fruit fly compound eye. B) Cartoon of the structural module of an ommatidium, where the photoreceptors visible in cross-section are coloured in red. C) Pseudopupil assay in wild type (left) and HD flies (right). [From Green and Giorgini 2012].

The fly life cycle is characterised by short reproductive and developmental cycles (10–14 days from embryo to reproductively mature adults) and a short life span (~12 weeks in wild type background), as well as an absence of meiotic recombination in males, all features that, with the easy handling and cheap maintenance, make the fly an amenable model organism with which to study many genetic disorders.

A wide range of genetic manipulation techniques have been developed within the fruit fly model to test the effect of overexpression, downregulation, deletion of endogenous and exogenous genes. Several null-allele mutant lines have been developed using chemically induced mutagenesis e.g., using ethyl methane sulfonate (EMS) or by insertion of a P element, an endogenous transposon that allows semi-random insertion mutagenesis with preferential insertion into gene promoter regions (Ambegaokar, Roy, Jackson 2010; Spradling and Rubin 1982). The P elements have been modified for several applications like chromosome engineering, gene tagging, and inducible gene expression/repression in combination with binary system of conditional expression. In fact, the expression of a gene of interest can be regulated using the GAL4/UAS system of expression, in which the yeast-derived GAL4 transcription factor binds to the upstream activating sequence (UAS) enhancer element, regulating the expression of the gene downstream of the UAS (Figure 14) (Brand and Perrimon 1993). The GAL4/UAS system offers the possibility to regulate the conditional expression of all mutant lines developed in fly: “enhancer–promoter” (EP) lines are easily available from large worldwide stock such as the Bloomington Drosophila Stock Center, http://flystocks.bio.indiana.edu/). EP construct consists in a P element carrying several GAL4 binding site within the target element enhancer to highly activate the gene of interest (Rørth 1996). Moreover, the tissue specific GAL4/UAS system allows the development of tissue specific knockdown of almost every gene within the fruit fly genome, by incorporating an antisense construct by RNA interference (RNAi) (collection of these lines is provided by the Vienna Drosophila RNAi Center, http://stockcenter.vdrc.at/control/main) (Ryder and Russell 2003). RNAi occurs through a UAS-construct carrying a target gene where is present an inverted repeat sequence (of at least 19 nucleotide). The transcript encoded by this sequence will be responsibile for the activation of the RNAi machinery with a resulting knock down of the gene of interest (Dietzl *et al.* 2007).

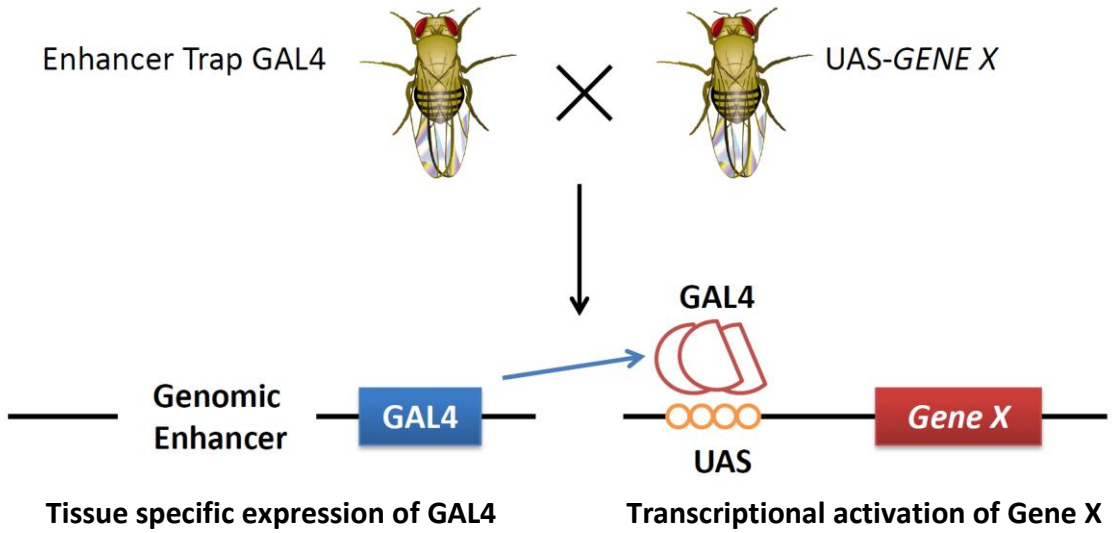


Figure 14. The binary GAL4/UAS system of gene expression in Drosophila.

The system GAL4/UAS allows the conditional expression of a gene of interest in a specific tissue, using GAL4 under control of tissue-specific enhancers, and the expression of deleterious gene only in experimental lines (allowing the maintenance of precursor lines). [Adapted from Brand and Perrimon 1993]

The simple genetics, the specialized nervous system and the available genetic tools make the fruit fly a good model system for in vivo neurodegenerative studies of HD. Different studies have already demonstrated that human pathogenic forms of disease genes produce relevant pathological phenotypes in flies (Kazemi-Esfarjani and Benzer 2000; Marsh *et al*. 2000; Nagai *et al*. 2003). Regarding HD, independent studies developed a number of fruit fly models of HD. These models express a wide repertoire of HD-associated changes evident from earliest developmental stages of the fly and, therefore, represent a good tool for the discovery of therapeutical targets. Although the *Drosophila* genome includes an orthologue of the human HTT, most of the fly HD models are based on the incorporation of an exon 1 toxic fragment from the human HTT. The majority of these models allow the study of conditional expression of the HTT construct, using the GAL4/UAS system utilising different GAL4 drivers such as elav>GAL4, which is routinely used and expresses UAS constructs pan-neuronally from an early stage of the fly life cycle (Green and Giorgini 2012).

Fly HD models show progressive and age-dependent neurodegenerative pathology directly correlated with the polyQ length (late in larval/pupal life), characterised by behavioural and locomotor dysfunctions and premature death. The neurodegeneration in HD flies is easily monitored by the analysis of the photoreceptors of the eye compound (pseudopupil assay), direct indication of the progressive neuronal loss (Figure 13) (Marsh, Pallos, Thompson 2003).

The application of *Drosophila* HD models has yielded significant insights into several aspects of HD pathogenesis, such as trafficking and axonal dysfunction (Gunawardena *et al*. 2003; Steinert *et al*. 2012). Fly HD studies identified several suppressors of polyQ toxicity, being as well an appropriate model organism for rapid screening of pharmacological compounds (Campesan *et al*. 2011; Green *et al*. 2012; Ravikumar *et al*. 2004; Steinert *et al*. 2012). Studies in regard of pharmacological interventions on the histone deacetylase (HDAC), have proven to be effective in fly and cell HD models, suggesting a potential therapy strategy against HD, which is now under preclinical investigation (Pallos *et al*. 2008; Steffan *et al*. 2001; Chen *et al*. 2012; Jia *et al*. 2012).

1.2.5.3 Rodent HD models

Predominantly based on the use of mouse lines, the rodent lines are most widely used system for modelling HD. Mouse and human share similar development and exhibit high genomic synteny, features that classify mouse models as eligible tools for studying HD, in addition to the availability of well-developed technologies for genome manipulation, screening and behavioural studies (Sosa, De Gasperi, Elder 2012).

All the rodent HD models produced are genetically engineered using different constructs within the full-length human HTT or a fragment containing the whole or partial human HTT exon 1. These constructs bear the HD mutagenic locus within between 80 and to 150 CAG repeat units, which ensures the transcription of the extended and pathogenic polyglutamine tail. The expression of the constructs can be regulated by either a human or an endogenous promoter according to the different models. Symptoms, life span and neuropathological features are quite variable for the several models (Zuccato, Valenza, Cattaneo 2010). Besides transgenic mice models of HD, also knock-out and knock-in mouse

models of HD has been generated. The knock-out models due to embryonic lethality allowed to identify the crucial role of HTT in embryogenesis, but are not widely used. The knock in mouse models bear the CAG mutation into their endogenous huntingtin gene (*Hdh*) or they carry a chimeric *HTT/Hdh* sequence, encoding for the polyglutamine stretch. Because of the appropriate genomic and protein context, the disease phenotype is milder and more similar to the human HD, making the knock-in mouse models faithful genetic models of the human condition (Menalled 2005).

Among the transgenic models, the R6/2 line, designed with the human *HTT* promoter followed by exon 1 within 144 CAG repeats, represents an efficient model of HD, widely used in the dissection of the pathology and for therapy screenings (Gil and Rego 2009; Mangiarini *et al*. 1996). The R6/2 line replicates well the human HD phenotype with early and progressive manifestation of motor and cognitive symptoms and reduced lifespan. As most of the mice models, neuronal inclusions and atrophy are evident in R6/2 brain, where neurotransmitter dysfunctions, reactive astrogliosis and aberrant synaptic plasticity are also detected.

1.2.6 Therapy in HD

An effective therapy for the cure of HD has not yet been developed. Only few compounds ameliorating primary symptoms as psychiatric disturbances and chorea are available (Reilmann 2013). However, several proposed therapeutic treatments are under experimental investigation. The strategies followed for the development of an HD therapy are mainly three. Commonly, potential treatments are designed to restore or control the altered functionality of interacting factors or pathways that are involved in the disease, with the aim of decreasing or delaying the progression of HD. A more recent approach focuses on mutant HTT with the intent of blocking the expression with RNA interference or interfering with the toxic polyQ affecting aggregation. Furthermore, still at an experimental level, a potential treatment of HD is based on stem cells transplantation in lesioned areas of the HD brain (Zuccato, Valenza, Cattaneo 2010).

So far, several potential therapeutic targets have been identified and many candidate drugs have been tested in different model organisms showing improvement of motor and/ or cognitive dysfunctions, but only one drug passed all the clinical phases of screening. Tetrabenazine (TBZ), an inhibitor of the dopamine pathway, is a the only therapeutical compound accepted by the Food and Drug Administration (United States) for the treatment of choreic symptoms in HD. TBZ, reducing monoamines and serotonin from pre-synaptic central nervous system neurons, alleviates the motor deficits and reduces striatal cell loss in HD mice. Human clinical trials confirmed that the use of TBZ reduces uncontrolled movements, albeit with several side effects such as dyspahgia, depression or parkinsonism (Frank *et al*. 2008; McLellan, Chalmers, Johnson 1974; Tang *et al*. 2007).

Obstacles in designing a therapy for HD are mainly based on the discrepancies between model organisms and humans. The failure of several drugs during clinical trials suggested that several mechanisms in human HD are different from its models and/or still unknown. Therefore, it is necessary to test single compounds in several HD models, bearing different levels of disease severity. Moreover, the high heterogeneity manifested in HD human individuals, correlated with the multi-factorial functionality of the HTT protein, suggests that an efficient therapeutical approach will be based on the combination of several compounds. This strategy could be valid if only the triggering pathogenic events, distinct from the several epiphenomena in the disease, will be targeted. For this reason it is necessary to identify presymptomatic biomarkers of the disease progression and, among the corresponding pathways, genetic factors with the power to modify the symptoms onset in HD. Considering that the identification of biomarkers is based on laborious screening (which are usually related to a small number of individuals due to technical and experimental duties), the discovery of genetic modifiers of the disease onset in large disease cohort studies and their following validation in different HD models represent a valid approach to identify new therapeutical targets.

1.3 Candidate genetic modifiers of AO in HD

CNVs as genetic variants are widespread in the human genome and with their potential to affect gene expression, it stands that they could be involved in HD pathology. A variance of the genomic copy number of a gene, implicated in any of the disease pathways, could result in either protective or toxic effects and be a genetic modifier of the AO in HD. CNV in the human β-defensin region and CNV involving the member 3 of the facilitated solute carrier family 2 (*SLC2A3*) locus, with a potential role respectively in neuroinflammation and the neuronal glucose uptake, were investigated in this study.

1.3.1 Human β defensin 2 (*DEFB4*)

Human β-defensin 2 (h-BD2) is a member of the defensin family, a category of antimicrobial peptides. Most defensins are small cationic peptides, with a molecular weight ranging from 2 to 6 kDa, able to interact with the bacterial membrane during the immune response in vertebrate and invertebrate organisms. In primates the defensins are characterized by a conserved motif of 6 cysteine residues (C-X6-C-X4-C-X9-C-X6-CC) that confers a peculiar tertiary structure; different disulphide bond configurations of this motif distinguish α, β and θ defensins (Lehrer 2004). Among all defensins, identified in several species, β-defensins are likely to be the common ancestor of all vertebrate defensins (Xiao Y *et al*. 2004).

All over the human genome there are three main clusters encoding for β-defensins located on chromosome 8, 6 and 20. hBD2 is encoded by the *DEFB4* gene, within a genomic region of ~2 kb, located in the defensin cluster on the chromosomal region 8p21-p23, which also includes several other β-defensin genes *DEFB1, DEFB103, DEFB104, DEFB105, DEFB106, DEFB107)*, and all the α-defensins genes. *DEFB4* genomic structure, typical of most of the β-

defensin genes, consists of two exons and one intron, encoding a precursor peptide and a post translational signal for the modification of this precursor into a mature peptide.

Isolated for the first time from lesional psoriatic skin, hBD2 (64 aa) is expressed at a basal level in epithelia (for example in the respiratory tract, gastrointestinal tract, urogenital system, pancreas and skin), and also in leukocytes and the bone marrow (Harder *et al*. 1997). However, its up-regulation can be induced in response to a wide range of pro-inflammatory stimuli including Gram-negative bacteria, lipopolysaccharides (LPSs), and interleukin-1α (IL-α), interleukin-1β (IL-1β), tumor necrosis factor-α (TNF-α), interferon-gamma (IFN-γ). These factors promote the expression of h-BD2 through the nuclear transcription factor κB (NF-κB), sometimes supported by activated protein 1, by mitogen-activated protein kinase (MAP-kinase) and by protein kinase C (PKC) (Pazgier *et al*. 2006). Following their activation, β-defensins "cross-talk" with the adaptive immune system via an interaction with toll-like receptors (TLRs), which recognize pathogen-associated molecular patterns and modulate the expression of several genes (Biragyn *et al*. 2002; Froy 2005; Weinberg *et al*. 2012; Yang *et al*. 1999).

h-BD2 is a potent immunomodulator with an important role in regulating both innate and adaptive immune responses. The activation of h-BD2 is most powerful in the presence Gram-negative bacteria and some fungi but less potent against Gram-positive bacteria, particularly at low concentrations of salt and plasma proteins ties. From an immunoregulatory perspective, h-BD2 shows chemoattractant activity, engaging a number of cell surface receptors such as the CCR6 receptor on immature dendritic cells (DC) and peripheral blood memory T cells and, in a chemokine manner, recruits these cells to the sites) of interest. This feature defines the important role of h-BD2 in signalling pathways with respect to the innate and adaptive immune response (Ganz 2003; Hardwick *et al*. 2012).

1.3.1.1 DEFB4 CNV

DEFB4 and other β-defensin genes on 8p23.1, with the exception of *DEFB1,* are on a large repeat unit that shows multiallelic CNV, as several studies using different CNV discovery tools have confirmed (Conrad *et al*. 2010; Hollox, Armour, Barber 2003). Normal

population have between 2 to 8 copies per diploid genome, but rare individuals with up to 12 copies have been identified by cytogenetic analysis (Hollox, Armour, Barber 2003; Linzmeier and Ganz 2005). In respect to CNV size (~250 kb), gene content and numeric variability of β-defensin region, 8p23.1 represents one of the CNV hotspot in the human genome (Figure 15) (Groth *et al*. 2010).

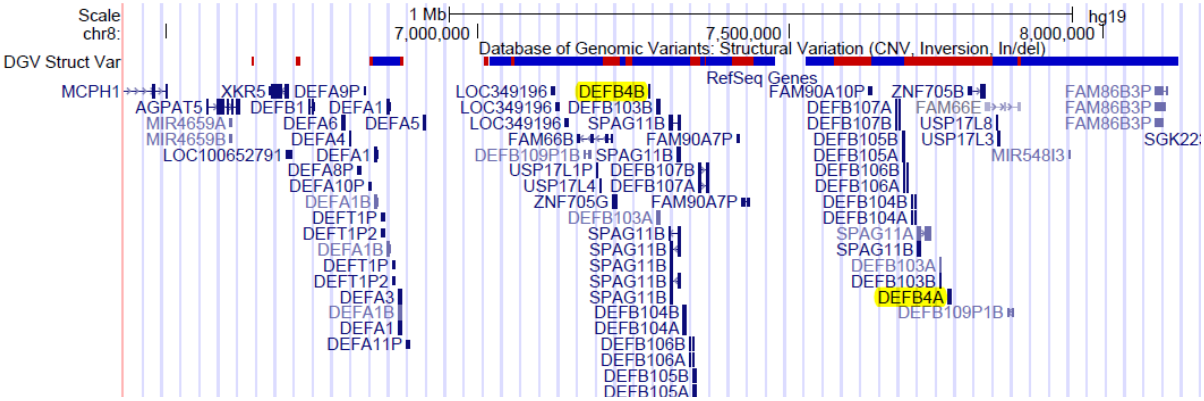


Figure 15. β-defensin region.

The chromosomal region is shown at the top of the diagram. DGV Struct Var bar refers to the Database of Genomics Variants from which Conrad et al. data are reported in the diagram (Conrad et al. 2010). The blue bar indicates a gain in size relative to the reference; the red bar indicates a loss in size relative to the reference. In the lower half of the diagram are indicated all the reference genes located in this region. Highlighted in yellow the 2 isoforms of DEFB4, DEFB4A and DEFB4B). [From (GRCh37/hg19) Assembly,Genome browser, http://genome.ucsc.edu].

Quantitative analysis of h-BD2 mRNA and protein levels in lymphoblastoid immortalized lymphocyte cell lines (LCLs), serum and cultured keratinocytes showed the direct correlation between *DEFB4* copy number and h-BD2 expression levels. (Groth *et al*. 2010; Jansen *et al*. 2009; Hollox, Armour, Barber 2003). Interestingly, higher genomic copy number of β-defensin region was associated to a greater risk of psoriasis, an inflammatory skin disease. Likely caused by a gene copy number-dosage effect, the higher h-BD2 protein level found in psoriatic skin tissue could be an early enhancer of inflammation (Hollox *et al*. 2007; Zeeuwen *et al*. 2008). Interestingly, higher β-defensin copy number was associated with increased HIV load and impaired immune reconstitution following initiation of highly active antiretroviral therapy in Sub-Saharan Africans individuals (Hardwick *et al*. 2012).

These studies suggest that β-defensin copy number variation may be a risk factor in other infectious and inflammatory diseases (Hardwick *et al*. 2012).

1.3.1.2 DEFB4 CNV: a potential genetic modifier of AO in HD

Studies performed in rats showed that the β-defensins r-BD1 and r-BD2, the latter which shares ~ 50% amino acid identity with h-BD2, are constitutively expressed in the brain (Froy *et al*. 2005). In human, the expression of h-BD2 in the brain was analysed in few studies (Hao *et al*. 2001; Tiszlavicz *et al*. 2011). One study tested the expression of h-BD2 and h-BD1 (founded previously in human brain biopsy tissue) in cultured astrocytes, microglia, meningeal fibroblasts and neurons, all derived from normal human (foetal) brain, in normal and inflammatory conditions. h-BD2 expression was inducible and only detected in astrocyte cultures, mainly after inoculation with LPS (within a dose sensitive effect) and IL-1β or TNF-α (Hao *et al*. 2001). Furthermore, it was found that brain capillary endothelial (BB19) cells, which are a human cell culture model for the blood-brain barrier (BBB), express hBD-2 upon infection of *Chlamydia pneumoniae* (common cause of infection and inflammation of the brain and meninges) (Tiszlavicz *et al*. 2011).

In common with most age-associated neurodegenerative disorders, neuroinflammation is a typical hallmark of HD (see paragraph 2.4.6) (Harry and Kraft 2008; Hirsch and Hunot 2009; Möller 2010; Rogers 2008). The number of activated microglia and reactive astrocytes were shown to be correlated with disease severity in mice and patients with HD, whereas the suppression of microglial activation prolonged the lifespan of HD mice (Hsiao *et al*. 2013; Pavese *et al*. 2006; Politis *et al*. 2011; Sapp *et al*. 2001; Tai *et al*. 2007; Zwilling *et al*. 2011). Moreover, the innate immune activation detectable in plasma of presymptomatic HD individuals suggests that an innate or adaptive response of the immune system is present in HD pathology (Björkqvist *et al*. 2008). Inflammatory changes in peripheral and CNS tissues could be independent and regulated by mutant HTT, ubiquitously expressed, causing analogous derangements centrally and peripherally; or an initial peripheral inflammatory response to the mutant HTT could spread to the CNS, through the passage of immunomodulatory elements across the BBB.

In the CNS, h-BD2, expressed in activated astrocytes and in BBB cell models, could play an important role in the neuroinflammation observed in HD (Figure 16).

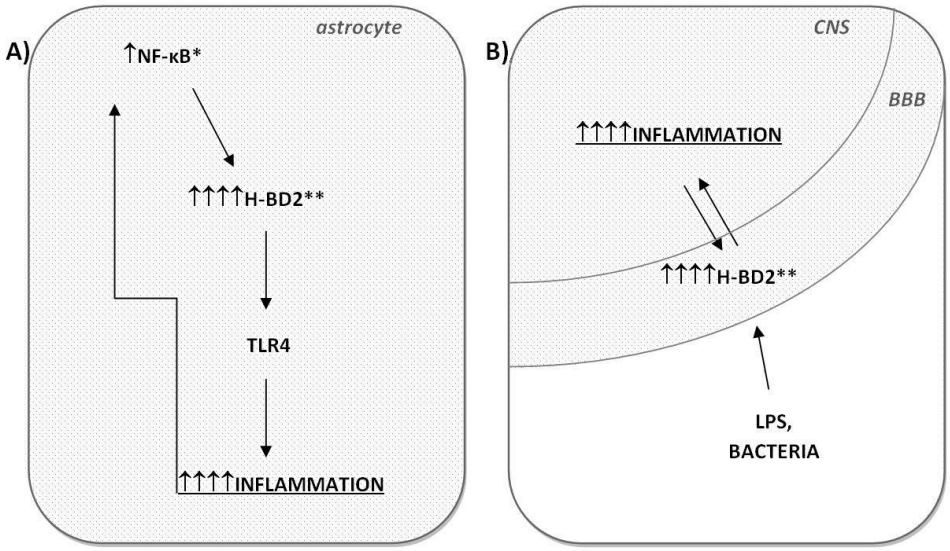


Figure 16. Proposed inflammatory role of h-BD2 in HD.

A) NF-KB expression in astrocytes activates h-BD2 in a copy number dependent way and h-BD2 through TLR4 mediates an exacerbation of neuroinflammation. B) Inflammatory stimuli passing through the BBB could activate h-BD2 expression in a copy number dependent way leading to an activation of neuroinflammation.

**Known to be highly expressed in HD. ** Expression dependent by DEFB4 copy number.*

The aberrant activation of NF-κB occurring only in HD reactive astrocytes could induce the expression of h-BD2, which in turn could exacerbate the inflammatory response interacting with toll receptor like 4, a key player in the neuroinflammation in AD and PD (Figure 16) (Biragyn *et al*. 2002; Hsiao *et al*. 2013; Fellner *et al*. 2013; Michaud *et al*. 2013). Moreover, h-BD2 expression could be induced by immuno-modulators through the BBB, leading to an increased neuroinflammation in HD (Figure 16).

h-BD2 could play an important role in HD neuroinflammation and β-defensins CNV could drastically change its relevance in the context of disease, where high copy number, and so high circulating protein level, could exacerbate inflammatory response mediated by h-BD2 and aggravate HD progression (Figure 16). Thus, β-defensin CNV could be a genetic modifier of AO in HD.

1.3.2 The neuronal glucose transporter GLUT3 (*SLC2A3*)

Glucose is the essential fuel to most mammalian cells; its passage across cell membranes is facilitated by a family of integral membrane transporter proteins, the GLUTs. Among these transporters there are the members of the facilitated solute carrier SLC2 family including the neuronal glucose transporter 3 GLUT3 (Kayano *et al*. 1988; Nagamatsu *et al*. 1992). Based on sequence, functional and predicted structural similarities the 14 GLUTs have been divided into three classes and GLUT3 together with GLUT1, -2, and -4 comprise the class 1 (Joost *et al*. 2002). These transporters differ in localization, hormone sensitivity and glucose facilitated transport. Interestingly, GLUT3, which is the main neuronal GLUT, has both a higher affinity for glucose than GLUT1, -2, or -4 and at least a five-fold greater transport capacity than GLUT1 and -4 (Simpson *et al*. 2008).

GLUT3 protein (45kDa) is encoded by the gene *SLC2A3,* within a genomic region of 17kb on chromosome 12p13.31, in the proximity of a highly homologous gene, *SLC2A14*, which encodes GLUT14 (mainly expressed in testis). With *SLC2A14*, *SLC2A3* is located in a highly repeated sequence genomic region where NAHR might occur, leading to CNV of between 1 to 3 copies of the two genes (Conrad *et al*. 2010). Although the X-ray crystal structure of GLUTs is not available, GLUT3 protein, like all the sugar transporter family members, is predicted to have twelve transmembrane domains that combined to form a central aqueous pore or channel through which the substrate crosses the lipid bilayer (Figure 17) (Simpson *et al*. 2008).

Expressed in placenta across early stage of gestation, GLUT3 has an important role for the energy exchange between embryo and mother and is essential during developmental stages (Brown *et al.* 2011). Mouse embryos homozygous for the *Slc2a3* null-allele (bearing a deletion of exons 7-10) showed increased apoptosis, developmental restriction and premature death at early stage of gestation (Ganguly *et al*. 2007). Heterozygotes for the null allele ($Slc2a3^{+/-}$) are viable but are associated with an abnormal distribution of GLUT1 in the blastocyst, suggesting the role of GLUT3 in polarization. In adult life, *Slc2a3* haploinsufficiency in mice does not affect brain glucose intake, body weight and growth pattern but is associated to a sexually dimorphic adiposity with insulin resistance in

males and features of autism spectrum disorders (Ganguly and Devaskar 2008; Stuart *et al*. 2011; Zhao *et al*. 2009).

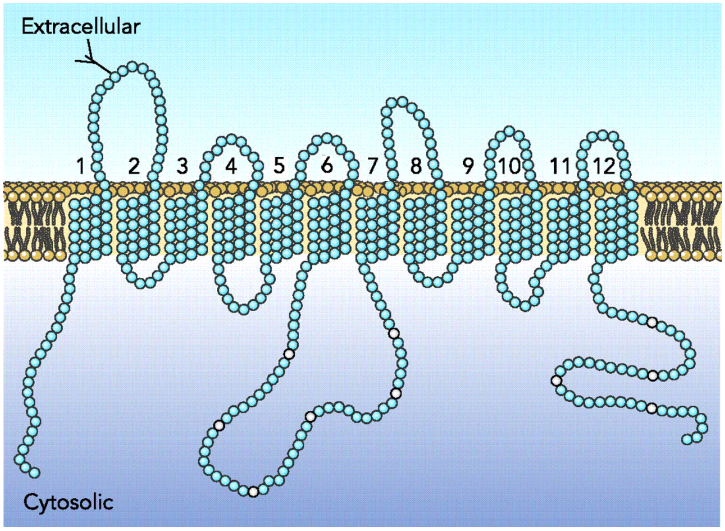


Figure 17. Putative structure of GLUTs of class I and II.

[Adapted from Manolescu et al. 2007].

Throughout cerebral maturation, GLUT3 mRNA and protein expression increases in a regional and activity-dependent manner; its expression level coincides with areas of the brain with high glucose demand, maturation and synaptic connectivity (Duelli and Kuschinsky 2001; Khan *et al*. 1999; Zeller *et al*. 1995). Among all the GLUTs present in the brain, GLUT3 is characterised by a high glucose capacity and within its localization is the main player of the glucose metabolism in the neurons. Physiological stimuli such as insulin, insulin growth factor 1 (IGF-1), and glucose deprivation can promote the translocation of GLUT3 to the plasma membrane but they do not influence the glucose uptake in neurons (Cheng *et al*. 2003; Uemura and Greenlee 2006). Furthermore GLUT3 expression can be sensitively altered during aging or in pathological conditions that affect the brain functionality and metabolism such as hypoxia after ischemic brain injury, AD and brain tumour (Boado, Black, Pardridge 1994; Simpson *et al*. 1994; Vannucci *et al*. 1998).

GLUT3 is also expressed in human white blood cells, namely in lymphocytes, monocytes/macrophages, neutrophils, and platelets. Its expression can be up-regulated following inflammatory stimuli in monocytes and lymphocytes, suggesting a role of GLUT3 in the immune response (Fu *et al*. 2004).

1.3.2.1 Cerebral glucose metabolism in HD

Aberrant metabolism of the CNS is a typical hallmark for several brain diseases. The glucose utilization rate is a valid age dependent parameter for tracking functional and metabolic change in the brain. Commonly detected by 18F-fluorodeoxyglucose (18FDG) PET technique, glucose utilization brain scans have been performed in the study of several brain diseases including depression, multiple sclerosis, and AD, contributing to a better understanding of these pathologies and being used as an efficient tool for the diagnosis and tracking of complex disorders such as AD (Baxter Jr *et al*. 1989; Mosconi 2005; Roelcke *et al*. 1997; Yoshii *et al*. 1988).

18FDG-PET scanning of HD carriers revealed the loss of the glucose uptake in the striatum (caudate and putamen) and cortex (frontal and temporal lobes), remarkably before the onset of clinical symptoms of the disease reference). In asymptomatic HD gene carriers the striatal metabolism is significantly decreased in absence of atrophy and the progression rate of HD shows a better correlation with the detected hypometabolism independent from the CAG length (Antonini *et al*. 1996; Ciarmiello *et al*. 2006; Grafton *et al*. 1990; Hayden *et al*. 1987; Kuhl *et al*. 1982; Mazziotta *et al*. 1987). Interestingly, a 5 year follow-up study on asymptomatic HD subjects showed that the caudate glucose metabolism was significantly decreased in subjects that became symptomatic in the course of the study, and this change was not correlated with the CAG mutation length (Ciarmiello *et al*. 2012). In the same study, the relative caudate glucose metabolism added to the CAG mutation length increases the regression coefficient in the prediction of the AO, suggesting the glucose metabolism as a good predictor of disease onset (Ciarmiello *et al*. 2012; Shin *et al*. 2013).

The role of the glucose metabolism in HD has also been demonstrated in several HD model studies. R6/2 mice showed an early and progressive metabolic impairment significantly associated with lower glucose uptake, independent of cell loss; primary cortical

neurons, derived from HD^{140Q} knock-in mice, showed a reduced uptake of glucose that was restored by reinforcing Rab11 function, altered in HD. Increasing glucose entry in glia was found to reduce glia-induced pathology in fly models of HD Besson *et al*. 2010; Cepeda-Prado *et al*. 2012; Li *et al*. 2012).

These studies suggest that dysfunction of glucose metabolism is actively contributing to the disease progression and its restoration is a potential therapeutic target in HD.

1.3.2.2 SLC2A3 CNV in HD: a potential genetic modifier of AO in HD

In the complex scenario of the possible causes leading to a reduced glucose metabolism in the HD brain, several studies focused on the role of insulin, whose release appears to be affected by mutant HTT expressed in pancreatic β-cells. Besides insulin dysfunction, a possible cause of the reduced glucose uptake in the brain might be caused by alteration of GLUT3, the glucose uptake mechanism that is insulin and IGF-1 independent. From a post-mortem brain study, a significant reduction of GLUT3 and GLUT1 protein expression levels was found in the caudate of HD brain at grade 1 and grade 3 but not in the cortex compared to controls samples. This suggests that the lessened expression of GLUT1 and GLUT3 transporters is specific for HD caudate and not merely a feature of *post-mortem* degradation or HD-specific brain atrophy (Gamberino and Brennan 1994). Interestingly, a decreased concentration of GLUT1 and GLUT3, not correlated to synaptic loss, was found as well in AD brain, where this depletion correlates with abnormal hyperphosphorylation of tau (Liu *et al*. 2008; Simpson *et al*. 1994). Furthermore, using the fruit fly model to study AD and HD it has been found that the increased expression of the fly ortholog of *SLC2A3* can suppress tau toxicity in the AD model and ameliorate locomotor activity and life span in a glial HD model (Besson *et al*. 2010; Shulman *et al*. 2011).

These studies suggest that HD progression could be affected by GLUT3 expression level and functionality as glucose uptake is mainly dependent in the neurons by this transporter. CNV involving *SLC2A3* could directly affect its protein expression level by gene-dosage effect and have an impact of HD. In fact, according to a previous study, *SLC2A3,* encoded from a sequence on chromosome 12p13.31, is located in a copy number variable region where two units of a tandem duplication recombine together through blocks of high

sequence similarity (up to 94%). Through a specific recombination point, indentified in the study, NAHR involves the two units arising to a combination of gain or loss of a segment of 129 kb DNA shared between the two units (Figure 18) (Veal, *et al*. 2013). As it has been shown in different studies, NAHR in 12p.13.31 is quite rare as the frequency of duplication event is between 5 to 7% and the frequency of deletion event is around 0.5 up to 2% in the populations analysed (Conrad *et al*. 2010; Veal, *et al*. 2013). Recently it was found an association with *SLC2A3* deletion and a substantial protection against Rheumatoid Arthritis andwas found in the context of attention-deficit/hyperactivity disorder-related psychopathology (Lesch *et al*. 2010). Increased *SLC2A3* copy number encoding higher protein level could improve the glucose uptake rate in neurons and protect the neuronal functionality in HD. Thus, *SLC2A3* CNV could be a potential genetic modifier of AO in HD.

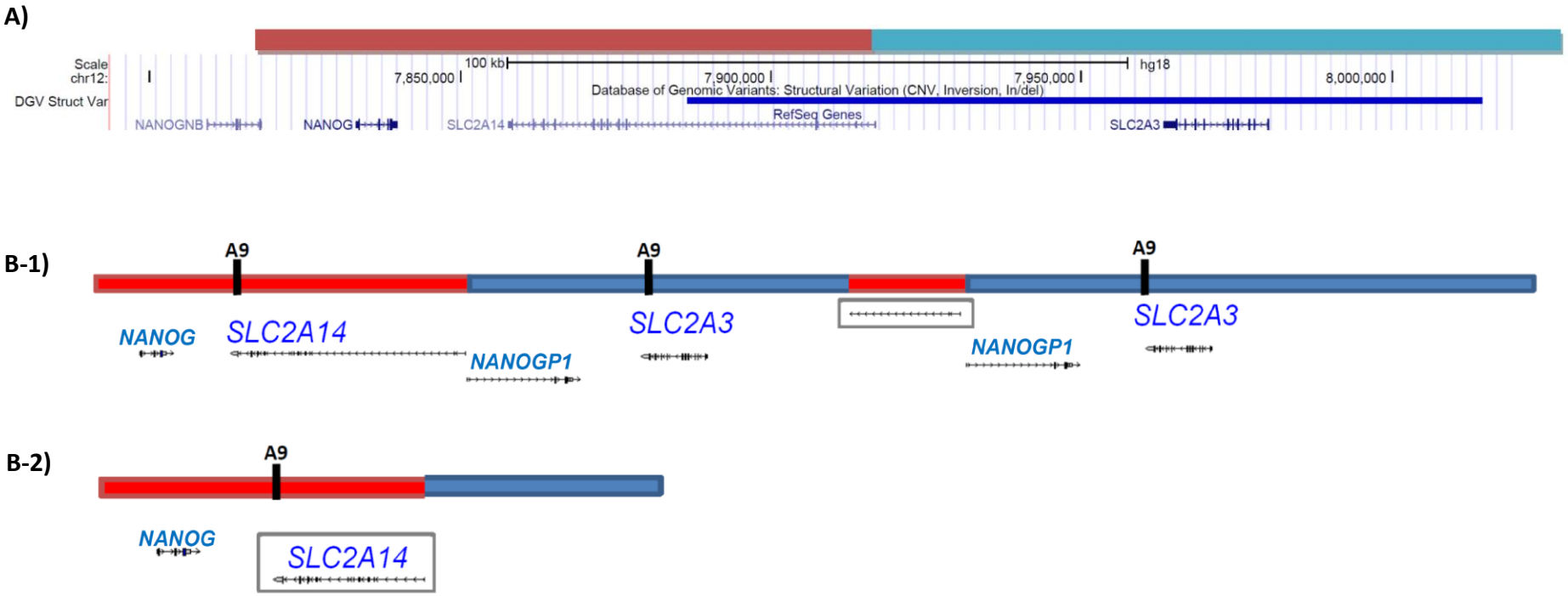


Figure 18A-B. Units of the tandem duplication involved in the NAHR that leads to CNV of *SLC2A3*.

A) The chromosomal region shown is part of chromosome 12p13.31 where SLC2A3 is located ((GRCh36/hg18) Assembly (http://genome.ucsc.edu)). DGV Struct Var bar refers to the Database of Genomics Variants from which Conrad et al., data are reported (Conrad et al. 2010). In the lower half of the diagram all the reference genes located in this region are indicated. In the upper part of the diagram the red and the blue bars represent the units of the tandem duplication involved in the NAHR for SLC2A3 CNV formation. B1,2) Hypotetical products of NAHR from a common recombination point in European ancestry individuals (adapted from Reekie 2011).The black lines indicated the annealing site for the primers of a SLC2A3 PRT assay (Veal et al. 2013). The grey boxes indicate fragments of the gene produced after the NAHR .B-1) Possible duplication product. B-2) Possible deletion product.

CHAPTER TWO: AIM OF THE STUDY

The aim of this study is to test *DEFB4* CNV and *SLC2A3* CNV as potential genetic modifiers of AO in HD. For this purpose we are going to analyse the distribution of diploid genomic copy number of these candidates in a cohort of HD individuals provided by the EHDN project, which includes only symptomatic subjects with variable AO and CAG mutation length, and is unbiased for ethnicity and gender. The study of CNV distribution for both candidates will be performed using the PRT approach, which allows rapid and efficient CNV genotyping of large data sets. Any positive association of our candidate genes with a variance of the AO in HD will require following validations in suitable *in vivo* HD models in order to test their functional role.

CHAPTER THREE: MATERIALS AND METHODS

Buffer compositions are given in section 3.7.

3.1 DNA samples

3.1.1 Control samples

All the DNA samples used as a standard controls were DNA samples purified from lymphoblastoid cell lines from the European Collection of Cell Cultures (ECACC), the HapMap panel, the Human Genome Diversity Cell Line (HGDP-CEPH) panel and the Leicester Panel. The controls sample list is reported in Table 25.

3.1.2 Disease cohort samples

Disease samples used in this study are unrelated European ancestry with HD, part of the EHDN "REGISTRY" project. REGISRTY enrols manifest and premanifest *HTT* mutation carriers, individuals at risk and controls into a multicenter observational study with annual follow-up visits. An informed written consent, according to the International Conference on Harmonisation-Good Clinical Practice guidelines was given by all the participants (http://www.ich.org/). Where there was inability of the participant to consent, the guidelines of the collecting country were followed. In case of minors, the assent of the parents was also requested. Furthermore, ethics approval was needed from the local ethics committee for each study site contributing to REGISTRY. Data collection uses electronic case report forms available in Czech, Danish, Dutch, English, Finnish, French, German, Italian, Norwegian, Polish, Portuguese, Spanish, Swedish and Russian. Each patient was assessed for motor, psychiatric and cognitive signs scored according to the Unified Huntington's Disease Rating Scale by the local clinicians. Blood samples was collected and shipped to BioRep (BioRep, Milano) in order to extract DNA, measure *HTT* CAG repeats length and for the creation of lymphoblastoid cell lines. Our study included 1000 HD manifest samples from REGISTRY, provided at different times in two separate cohorts of 500 samples each. The DNA samples were supplied in 96-multi well plate at different concentrations, which were adjusted with MilliQ water (Merck Millipore, Billerica, MA, USA) to the concentration of 10ng/µl. A nine digit pseudonym for each DNA sample was provided, allowing the identification of the samples in

the EHDN database. At the beginning of the study no clinical information was given in order to assess blind study on these cohorts.

3.2 Standard Methods

3.2.1 Polymerase Chain Reaction (PCR)

3.2.1.1 PCR in 10X KAPA Biosystem buffer A

The DNA sequence of interest was amplified by the PCR method. PCR mix using KAPA Biosystem 10X buffer A (15 mM Mg^{2+}) was commonly prepared in this study. Usually, 10µl PCR reactions were prepared as a master mix with the final concentrations of 1X buffer A, 0.2 mM each dNTP, 0.05U *Taq* DNA polymerase and 10ng input DNA. PCRs were performed in a Veriti thermal cycler as follows: 98˚C for 1 minute; 25-30 cycles of 95˚C for 15 seconds, annealing phase and 72˚C for 1 minute followed by a final extension of 72˚C for 5 minutes. Exact cycle temperatures and times depended on the primers and the expected product. For DNA sequence of interest where the CG base content was significantly high (CG>60% of total sequence), betaine was added to the mix at a final concentration of 0.1M.

3.2.1.2 PCR in 10X Low dNTPs (LD) PCR buffer

PCR mix using 10X Low dNTPs (LD) PCR buffer (3.7) was commonly prepared in this study. Usually, 10µl PCR reactions were prepared as a master mix with the final concentrations of 1X LD PCR buffer, 0.5U *Taq* DNA polymerase and 10ng input DNA. The primers were used at a final concentration optimized for each protocol. PCRs were performed in DNA Engine Tetrad thermal cycler (MJ research, Quebec, Canada) as follows: 98˚C for 1minute; 30 cycles of 95˚C for 30 seconds, annealing phase and 70˚C for 30 seconds, followed by a single “chase” phase of annealing for 1 minute and 70˚C for 20 minutes to reduce levels of single-stranded DNA and complete terminal 3’ dA addition. Precise cycle temperatures and times depended on the primers and the expected product length.

3.2.2 DNA electrophoresis

3.2.2.1 Agarose gel electrophoresis

To measure the yield and size of PCR products amplified, DNA was separated by non-denaturing agarose gel electrophoresis. Agarose (Lonza, Basel, Switzerland) was dissolved by boiling in an appropriate amount of 0.5X TBE (3.7) containing 0.5µg/ml ethidium bromides to give a gel in the range of 0.7-2% w/v. DNA samples and DNA ladders to be run were prepared with an appropriate amount of loading buffer and then loaded into the wells. Samples were run at 120V for 1 to 2 hours; bands were visualized by illumination under UV light and a photograph of the gel was kept for records using a Gene Flash Syngene Bioimaging device (Synoptics Inc, Frederick, MD, USA).

3.2.2.2 Capillary electrophoresis

Capillary electrophoresis was carried out in this study using ABI Genetic Analyzer 3130 XL instrumentation. GeneScan analysis was performed using fluorescently labeled DNA. 0.01 to 1µl of PCR products were mixed with 10µl HiDi formamide and 1% of the internal size standard GeneScan-ROX400 was included for precise determination of the length of the amplicons. After denaturation for 3 minutes at 98°C, the products were separated on POP-7 polymer with 30 seconds of injection time and collected by Genetic Analyzer data collection software v3.0. Scanning results of fluorescent-dye-labeled PCR products (peak area) by GeneScan v3.7 software were collected and transferred to an Excel spreadsheet file through GeneMapper v4.0 software for further analysis.

3.3 PRT-based assays

3.3.1 *DEFB4* assay

The *DEFB4* assay is a method developed in a previous study to measure the copy number of β-defensin per diploid genome (Aldhous *et al*. 2010). The assay consists of a duplex PRTs assay and one indel measurement assay, respectively called PRT107A, HSPD21 and 5DELR4. The assays

were carried out as follows (3.3.1.1) separately and analyzed all together in a single capillary electrophoresis run (3.2.2.2). The data from the duplex PRTs assay were validated as described in 3.3.1.1 and the copy number call was inferred from the PRTs and indel estimates by a maximum likelihood analysis, described in 3.3.1.3.

3.3.1.1 Duplex PRT assays

In order to measure the copy numbers of β-defensin per diploid genome, two different systems of PRT assays were performed following the protocol used in Aldhous *et al*. 2010 study. PRT107A amplifies a reference region on chromosome 11 ((chr11:97,384,722-97,384,877) (Febr.2009 (NCBI37/hg19) Assembly)) and a test locus on chromosome 8 ((chr8:7,387,448+7,387,605; chr8:7,648,479-7,648,636) (Febr.2009 (NCBI37/hg19)). This PCR is performed in duplicate for each sample using FAM- or HEX-labeled forward primer and an unlabelled reverse primer at the final concentration of 0.5μM (Table 26). HSPD21 amplifies a reference region on chromosome 21 ((chr21:30,260,100-30,260,279) (Febr.2009 (NCBI37/hg19)) and the test locus found on each of the β-defensin clusters on chromosome 8 ((chr8:7277405+7277576), (chr8:7749011-7749182) (Mar. 2006 (NCBI36/hg19) Assembly)). This PCR is performed in duplicate for each sample using an unlabeled forward primer and a FAM- or HEX-labeled reverse primer at the final concentration of 0.5μM (Table 26). HSPD21 and PRT107A products were amplified simultaneously using the same PCR mix with 10X LD PCR buffer (3.2.1.2) with pre-denaturation of 95˚C for 5 minutes, followed by 22 cycles of 95˚C for 30 seconds, 58˚C for 30 seconds and 70˚C for 1 minute, ending with a single "chase" phase of 58˚C for 30 seconds and 70˚C for 40 minutes. In total, two parallel amplifications were carried out for each sample, one with a FAM label, the other with HEX/NED; 1μl from each reaction was analyzed by capillary electrophoresis (3.2.2.2).

Data analysis

For PRT107A areas corresponding to the 158bp from near *DEFB107* and the 155bp from chromosome 11 and peak areas corresponding to the 172bp from near *DEFB4* and the 180bp from chromosome 21 in HSPD21 were recorded using Genetic Analyzer data collection software v3.0 and Genemapper v4.0 software. The ratio 158bp/155bp and 172bp/180bp for PRT107A and HSPD21 respectively were compared between FAM- and HEX- or NED- labeled products and the

results were accepted if the difference between the ratios was less than 15% of their mean; this criterion led to the rejection of about less than 20% of tests. If accepted, the mean of the FAM and HEX or FAM and NED ratios was used in further analysis. Mean ratios were used in conjunction with reference samples of known copy number to calibrate each experiment, and the resulting (least-squares) linear regression used to infer the copy numbers for unknown samples. Selected DNA samples (Table 25) giving reproducible results from several PRT assays tested in previous studies were used as calibration standards (Figure 19) (Aldhous *et al*. 2010; Fode *et al*. 2011).

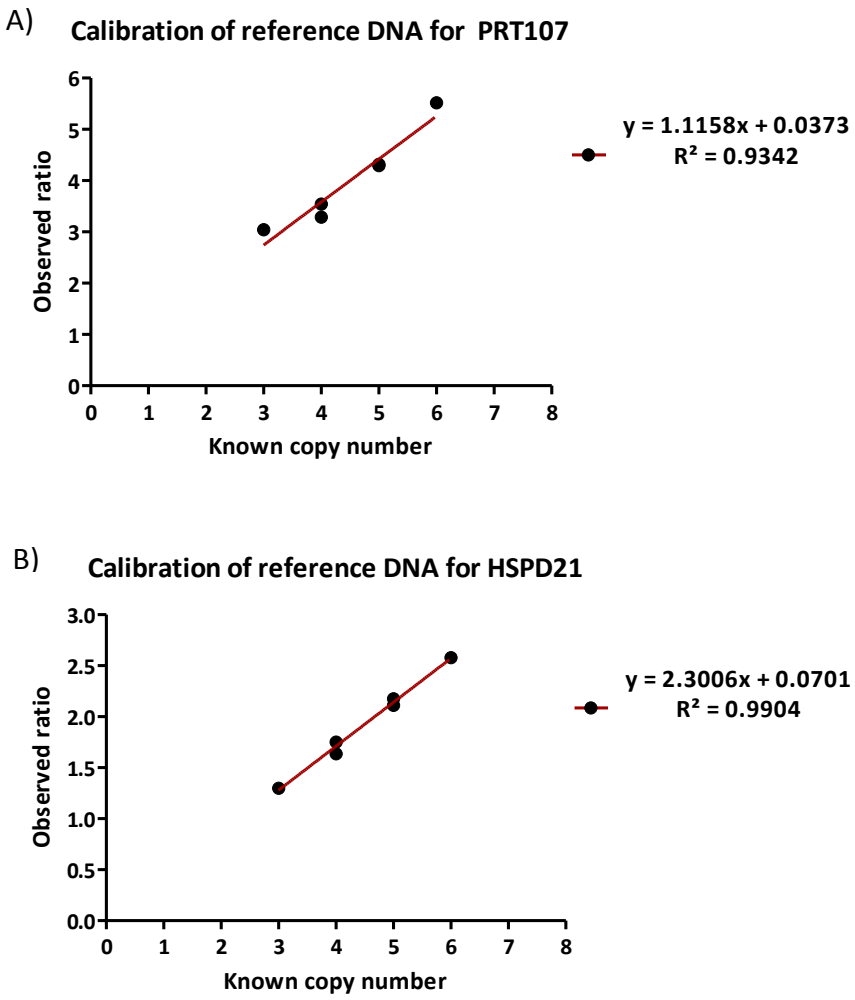


Figure 19 A, B. Reference DNA standard calibration for PRT107A (A) and HSPD21 (B).

An example of the calibration standard on selected reference DNA samples, which give reproducible results from the PRT systems. The linear regression (red line) shown above was used to infer the copy numbers for unknown samples.

3.3.1.2 Indel ratio measurements

In order to have further clarification about of the copy numbers of β-defensin in diploid genomes indicated by PRT assays, an indel ratio measurement assay was performed according to 5DELR4 assay from a previous study (Abu Bakar, Hollox, Armour 2009). 5DELR4 assay used the same couple of primers (Table 26) to amplify two separate regions at chromosome 8 ((chr8:7,659,491+7,659,624 and chr8:7,376,438-7,376,566 (Febr.2009 (NCBI37/hg19)) wherein rs71230882 is located. Depending on the triallelic indel polymorphism the products were variable in size (Table 26) and could differ for each *DEFB4* repeats. The PCR was performed using FAM- or HEX-labeled forward primer and unlabeled reverse primer at the final concentration of 5μM and the PCR mix with 10X LD PCR buffer (3.2.1.2). PCR amplification was performed as follows: 22 cycles of 95˚C for 30 seconds, annealing temperature at 50˚C for 30 seconds and 70˚C for 30 seconds, followed by a single "chase" phase of 58˚C for 1 minute and 70˚C for 20 minutes to complete 3' dA addition. Two parallel amplifications were carried out for each sample, one with a FAM label and the other with HEX; then 1μl from each reaction was added to 10μl HiDi formamide with ROX-400 marker and products and analyzed by capillary electrophoresis (3.2.2.2). All peak areas corresponding to the 5DELR4 respectively were recorded for both FAM- and HEX- labelled products using Genetic Analyzer data collection software v3.0 and Genemapper v4.0 software.

3.3.1.3 Maximum Likelihood Analysis

The Maximum Likelihood (ML) approach was used to infer the copy number call from the analyses combined together, namely PRT104, HSPD21 and 5DELR4 (Aldhous *et al*. 2010). The ML analysis calculates the probability that the observed estimate from each assays with error following a Gaussian normal distribution reflected each integer value from 1 to 9. For each sample the relative likelihood of the PRTs observed values referring to an integer value of 1 to 9 were calculated. The relative likelihoods of 5DELR4 observed ratio were calculated for integers of 1 to 9. Whereas the 5DELR4 assay gave three peaks, two ratios were estimated and the sum of the two ratios was calculated. The ML estimate of copy number for each sample was calculated multiplying the highest probability for each assay for each observed value. The standard deviation of the Gaussian distribution was estimated by multiplying the observed value by the coefficient of variation used as a threshold in the PRT or the maximum coefficient of variation from

experimental repeat measurements for 5DELR4 analysis. The level of confidence in each diplotype was calculated by -2(LN*b*-LN*a*) where *a* is the estimated probability for a specific copy number and *b* is the pooled probability for all other copy number diplotypes. Considering that this statistic follows a χ2 distribution, the p-value was determined using χ2 test for each copy number diplotype call.

3.3.2 *SLC2A3* copy number assay

In order to estimate the copy number change for *SLC2A3* locus, we performed a PRT assay, developed in a previous study (Veal *et al.* 2013), adapted to analysis on capillary electrophoresis analyzer, instead of agarose gel analysis. The primers were designed to amplify two paralogue genomic loci involved in their communal NAHR, the first lying on the last exon of *SLC2A3* ((chr12:8073299+8073583 (Febr.2009 (NCBI37/hg19)) and the second on the 3'UTR of *SLC2A14* ((chr12:8,073,299-8,073,583 (Febr.2009 (NCBI37/hg19)) (Table 26). Two parallel PCRs were set up using 10ng of DNA, 0.15 μM FAM and HEX-labelled forward primer, 0.15μM unlabelled reverse primer (Table 26) and a PCR mix following the PCR in 10X KAPA Biosystem buffer A protocol (3.2.1.1). Products were denatured at 98°C for 1 minute and then amplified using 25 cycles of 98°C for 15 seconds, 50°C for 15 seconds and 72°C for 1 minute and 70°C for 5 minutes to enhance complete extension in the final round and hence reduce levels of single-stranded DNA products. PCR products were analyzed by capillary electrophoresis (3.2.2.2).

Data analysis

The signal intensity of each product was calculated from the area of the peak detected by Genemapper v4.0 after capillary electrophoresis run. The ratio of product from each unit of the tandem duplication was calculated by dividing the peak area of *SLC2A3* amplicon by the peak area of *SLC2A14* amplicon. Mean ratios were used in conjunction with reference samples of known copy number (Table 25) to calibrate each experiment and the resulting linear regression was used to infer the copy numbers for unknown samples. The normalized ratio 285bp/200bp was compared between FAM- and HEX-labeled products and the results were accepted if the difference between the ratios was less than 15% of their mean; this criterion led to the rejection of less than 10% of tests. The mean of the FAM and HEX ratios was used to infer the copy number of the three different genotypic variants.

3.4 Statistical analysis of genetic modifiers of AO in HD

We analyzed the potential role of our candidates as genetic modifiers of the AO in HD testing their CNV frequencies in our cohort using SPSS 20.0 (IBM Inc., Armonk, NY, USA). The analysis was carried out creating a generalized linear model in analysis of variance and covariance. Firstly, we applied a model of analysis of variance with the expanded CAG repeat length as a scalar predictive variable, CNV as an ordinal predictor variable and the log2 transformed AO of HD as the scalar dependent variable, resulting in the best goodness-of-fit, estimated by the Wald Chi-Square ($\chi 2$). We then applied the same model to investigate any possible association of CNV frequencies and the major estimated symptoms at the onset of HD, using the CNVs and the expanded CAG repeat length as independent variables and the different symptoms as categorical dependent variables.

3.5 Lymphoblastoid cell lines from HD patients study

3.5.1 Cell lines used

The lymphoblastoid cell lines (LCLs) used were provided by EHND. The LCLs were originated from individuals of our HD cohorts (3.1.2). Five LCLs were selected among each group of individuals, previously genotyped, carrying 1, 2 and 3 copy number of *SLC2A3*. The cell lines were cultured in filter cap flask, 50 and 250 ml (Greiner Bio-One Ltd, Stonehouse, UK), at 0.5-1x10^6cell/ml concentration and cultured with RPMI 1640, GlutaMAX (TM) media, supplemented with 10% fetal bovine serum (Biowest SAS, Nuaillé, France) and 100 units/ml penicillin and 100μg/ml streptomycin. The cells were harvested at 37°C, in 5% of CO_2.

3.5.2 GLUT3 quantification by immunoblot

3.5.2.1 Protein extraction

Cells were counted using a haemocytometer and 10 to 20x10^6 cells were taken for protein extraction. The cells were centrifuged at 300*g* for 5 minutes and washed in Phosphate buffered saline (PBS). Cells were lysed for 10 minutes on ice, using a Lyses buffer (LB) (3.7). The volume of LB was adjusted according the cell concentration using 100µl of LB for each 10x10^6 cells. The lysates were centrifuged at 13,000 rpm for 10 minutes at 4°C and the supernatant was collected and stored at -80°C. In order to measure the protein concentration, 1µl of the protein supernatants was diluted in 1ml of 1:5 Bradford's Reagent (Sigma-Aldrich, St. Louis, Missouri, USA) dilution and consequently quantified by measurement of the optical density at FLUOstar Omega plate reader (BMG LABTECH GmbH, Ortenberg, Germany).

3.5.2.2 Sodium dodecyl sulphat-polyacrylaminde gel electrophoresis (SDS-PAGE)

An appropriate amount of 1X western blot (WB) loading buffer (3.7) was added to the 10µg of protein product from each sample in order to adjust the final concentration for the electrophoresis run and everything was incubated at 95˚C for 10 minutes. Two pre-stained broad range ladder (250-10 kD) were loaded into the gel with the samples. The stacking gel and the resolving gel, namely the two partition of the acrylamide gel, were prepared using the component respectively listed in Table 4 and Table 5. The gel was run using 1X running buffer (3.7) at 15mA for approximately 2 hours.

Table 4. Components list for stacking gel.

Component (Stock)	v/v in ~4ml
Stacking buffer*	1000 µl
30% Acrylamide**	680 µl
H_2O	2275 µl
Ammonioum persulfate(H_2O) 10%	40 µl
TEMED	4 µl

**see 3.7*

*** Acrylamide/Bis-acrylamide=37:1*

Table 5. Components list for 10% resolving gel.

Component (Stock)	v/v in ~10ml
Resolving buffer*	2.5 ml
30% Acrylamide**	3.3 ml
H_2O	4.0 ml
Ammonioum persulfate (H_2O) 10%	75 µl
TEMED	7.5 µl

**see 3.7*

*** Acrylamide/Bis-acrylamide=37:1*

3.5.2.3 Transfer onto Nitrocellulose Membrane

After the separation on the acrilamide gel, the proteins were electrically transferred onto polyvinylidene difluoride membrane. Each blotting sandwich was prepared surrounding the gel and the membrane either side by 3 layers of absorbent filter papers (Whatman plc, Maidstone, UK) to ameliorate the absorbance of the transfer buffer (3.7) used in this step. The protein blotting was done at 100 V for approximately 2 hours at 4°C.

3.5.2.4 Immunostaining with Antibodies and Protein Detection

Blocking buffer (3.7) with 5% w/v milk powder was used to block aspecific sites and proteins onto the nitrocellulose membrane for 1 hour. The membrane was then probed with 1:400 dilution of a rabbit polyclonal antibody against GLUT3 (ab15311, Abcam plc, Cambridge, UK) in TBS-T with 1%w/v milk (11.2) at 4°C overnight with continuous slight agitation. Subsequently the membrane was washed using TBS-T 3 times for 15 minutes each and with TBS-T with 1% w/v milk 2 times for 10 minutes. Afterwards the membrane was probed with 1:10,000 dilution of anti-rabbit IgG secondary antibody conjugated to horseradish-peroxidase (PI-1000, Vector Laboratories, inc., Burlingame, California, USA), in TBS-T with 1%w/v milk for 1 hour at 4°C with continuous agitation. The procedure was repeated for α-Tubulin, the reference protein, using for the immunostaning of the primary antibody a 1:200 dilution of a mouse antibody against α-

Tubulin (sc-8035, Santa Cruz Biotechnologies inc., Santa Cruz, California, USA) and for the secondary antibody a 1:10,000 dilution of anti-mouse IgG secondary antibody conjugated to horseradish-peroxidase (PI-2000, Vector Laboratories, inc., Burlingame, California, USA), both in TBS-T with 1%w/v milk. The proteins were detected through colorimetric analysis. The blot was treated for 1 minute with a solution containing ECL substrate (SuperSignal West Dura Extended Duration Substrate) (Thermo Fisher Scientific, Waltham, USA) and then exposed for an appropriate time to photographic film (Fujifilm, Minato, Japan), subsequently developed. Relative proteins levels were estimated by densitometry and mixed effects linear regression analysis was performed to study the variance in relative protein expression levels of GLUT3 in our samples.

3.6 Studies with HD model fruit-flies

3.6.1 Fly stocks

All the experiments were done using *D. melanogaster* adult flies grown on standard sucrose-yeast medium (72 g/L maize meal, 80 g/L glucose, 50 g/L brewer's yeast, 8.5 g/L agar , 2 g/L of Nipagine [methyl p-hydroxybenzoate], dissolved in 10 ml 100% ethanol). All stocks were reared either at 18°C or 25°C under a cycle of 12 hours of light and dark (LD 12:12).

The *D. melanogaster* strains used are described below:

- w^{1118}*;; :* strain carries a null mutant in the *white* gene that produces white colored eyes. This strain is used as a standard genetic background for the production of transgenic flies and in general is used as a control in experiments;

- *FM7a; Cyo/Sco;* and *FM7a;; MKRS/TM6B*: strains carry chromosome balancers respectively for the sexual chromosome, the second and the third chromosomes, associated with visible markers that are useful and allow following genes in crosses scheme;

- *y,w*;; *actinGAL4*/*TM6*: a yellow white strain carrying actin, an ubiquitous promoter which drives the expression of the yeast *S. cerevisiae* protein GAL4;

- *elavGAL4*;; : strain carries a P-element on the X chromosome that, fused with *GAL4*, drives expression of *UAS*-genes pan-neuronally;

- *w;; UAS-Q93httexon1*: strain carries a *UAS*-element on the third chromosome which permits the expression of the first exon fragment of human Huntingtin with 93 poly glutamine, upon *GAL4* activation (Steffan *et al*. 2001);

- *w*;; $Glut1^{17J}$ /*TM6B*: strain carries a lethal mutation caused by alkylating agent in *Glut1* gene (Wei *et al.* 2003);

- *w*;; $Glut1^{d05758}$: strain carries a transposable P-element inserted in the position 3L:959,584..959,584 of *Drosophila* genome (Fly base genome browser, http://flybase.org/cgi-bin/gbrowse/dmel/);

- *w*; $Glut1^{KK108683}$; : strain carries a construct designed to generate dsRNA for RNA interference (RNAi) specific for *Glut1* and inserted in a attP landing site (Vienna *Drosophila* RNAi Centre,Vienna, Austria);

- *w*; *KK*; : strain carries a construct designed to generate dsRNA for RNAi with no specific target and inserted in a attP landing site (Vienna *Drosophila* RNAi Centre (VDRC),Vienna, Austria);

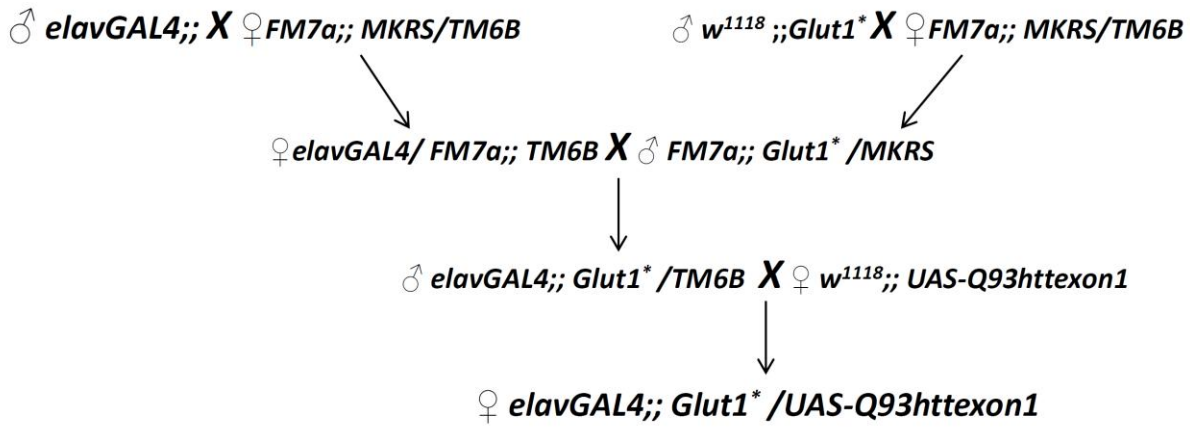
- *w*; *3M*; : strain carries an empty vector in the attP landing site used by VDRC, in general used as a control for experiment with RNAi construct.

Male flies carrying the tissue specific *GAL4* drivers were crossed to virgin females carrying the *UAS* construct(s) in order to obtain female in the F1 progeny expressing the transgene of interest.

3.6.2 Crossing scheme

In this study *elavGAL4* was the only driver used in order to induce the expression of the transgenic constructs of our interest in the *Drosophila* CNS. We investigated the effect of *Glut1* overexpression in HD background analysing *elavGAL4*;; $Glut1^{d05758}$/*UAS-Q93httexon1* for several metrics. This strain was obtained following the crossing scheme in Figure 20. Hereafter, we will refer to *elavGAL4*;; $Glut1^{d05758}$/*UAS-Q93httexon1* as EP-*Glut1*_Htt93Q. We investigated the effect

of single nucleotide mutation in *Glut1* in HD background analysing *elavGAL4*;; *Glut1*^{17J} */UAS-Q93httexon1* for several metrics. We investigated the effect of single nucleotide mutation in *Glut1* in HD background analysing *elavGAL4*;; *Glut1*^{17J} */UAS-Q93httexon1* for several metrics. This strain was obtained following the crossing scheme in Figure 20 and we will refer to it as mut-*Glut1*_Htt93Q.



"Glut1" refers to "Glut1*^{17J} */TM6B" or "Glut1*^{d05758}*".*

Figure 20. Crossing scheme adopted to generate EP-*Glut1*_Htt93Q flies and mut-*Glut1*_Htt93Q flies.

In this study EP-*Glut1* flies and mut-*Glut1* flies were compared to *elavGAL4*;; *UAS-Q93httexon1*, (abbreviated with Htt93Q). Control strains for EP-*Glut1*_Htt93Q, mut-*Glut1*_Htt93Q and Htt93Q (Table 27) were always analysed for each metric of investigation.

We explored the effect of *Glut1* downregulation by RNAi in HD background analysing *elavGAL4*; *Glut1*^{KK108683} ; *UAS-Q93httexon1* flies for several metrics. This strain was obtained following the crossing scheme in Figure 21. From now on we will refer to *elavGAL4; Glut1*^{KK108683} *; UAS-Q93httexon1* as *Glut1*-RNAi_Htt93Q. In this study *Glut1*-RNAi flies were compared to *elavGAL4*; 3M; *UAS-Q93httexon1* strain, where 3M refers to the insertion site designed for KK constructs for RNAi of *Glut1* (VDRC, Vienna, Austria) filled with an empty vector. From now on we will refer to *elavGAL4*; 3M; *UAS-Q93httexon1* as Htt93Q^{3M}. Control strains for *Glut1*RNAi_Htt93Q, and Htt93Q^{3M} (Table 27) were always analysed for each metrics of investigation.

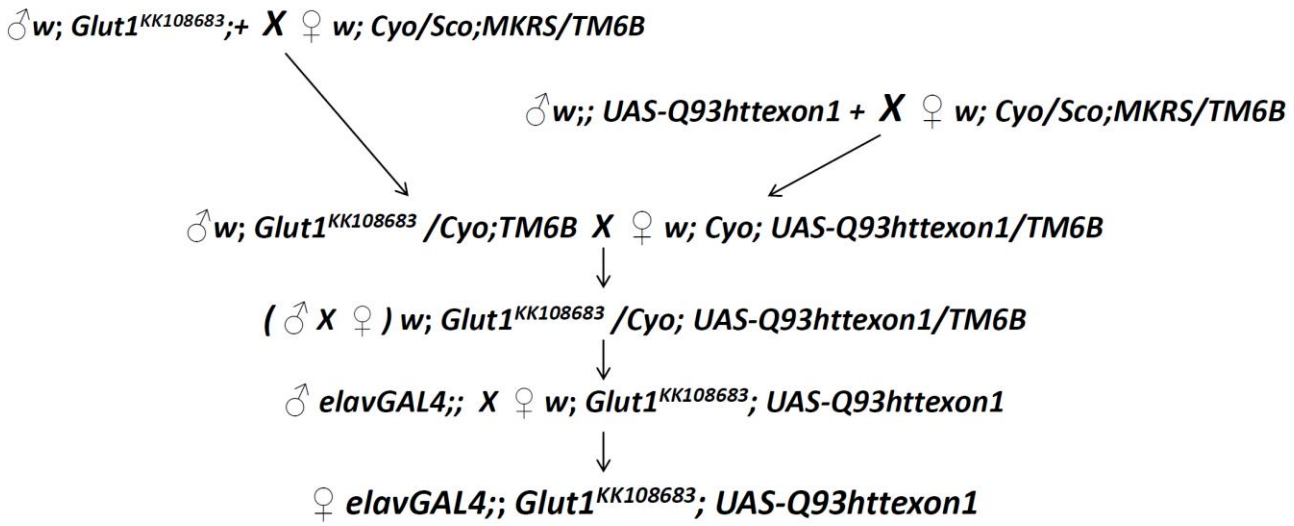


Figure 21. Crossing scheme adopted to generate RNAi-Glut1_Htt93Q flies.

3.6.3 RNA extraction

D. melanogaster fly samples (more than 15 samples for each strain) were collected at a specific time points during their lifespan, frozen in liquid nitrogen and stored in -80˚C fridge. Each batch of flies was homogenised in 400µl TRIZOL reagent using a plastic pastel. Ulterior 600µl of TRIZOL were added to the homogenate for 5 minutes at room temperature. Consequently, 200µl of chloroform were added to the mixture, which was shaken vigorously for 15 seconds. The mixture was left to stand for 10 minutes and then centrifuged at 13,000 rpm for 15 minutes at 4°C. The supernatant was collected and transferred into a RNA-free tube, 500µl of isopropanol were added to the solution, left to stand at room temperature for 10 minutes. To facilitate the collection of the RNA, the solution was centrifuge at 13,000 rpm for 15 minutes at 4°C. The supernatant was discharged and the pellet, containing the RNA precipitated, was washed in 80% EtOh and then re-suspended in 20µl of Diethylpyrocarbonate (DEPC)-treated dH_2O. The RNA extracted was tested for quality at 2100 BioAnalyzer and concentration at NanoDrop™ (Thermo Fisher Scientific, Waltham, USA).

3.6.4 cDNA synthesis

Before cDNA synthesis, RNA was treated with TURBO DNA-free™ kit to remove contaminating DNA, and afterward to remove the DNase and divalent cations from the samples. The treatment consisted in the preparation of this mix in a RNase-free microcentrifuge tube: 1-2µg of RNA sample; 1µl of 10X DNaseI reaction Buffer; 1µl DNase I, Amp Grade (1 U/µl); 10µl DEPC-treated H_2O. The reaction was carried out for 30 minutes at 37°C. The DNase was inactivated adding 2 µl of DNase Inactivation Reagent to the reaction mixture for 5 minutes at room temperature. The tubes were then centrifuged at 10,000 × g for 1.5 minutes and transfer the RNA to a fresh tube. cDNA synthesis was carried out using ImProm-II™ Reverse Transcriptase kit. 5µl of DNase treated RNA were mixed with 0.5µl of oligo-dT primer (100ng) and 0.5µl of Random Primers (100ng) at 72°C for 5 minutes and then chilled on ice for 5 minutes. The following components were added to the solution: 4µl of 5X ImProm-II™ Buffer; 2.5µl of 25mM $MgCl_2$; 1µl of dNTP mix (10mM of each); 0.5µl of Rnasin Ribonulcease Inhibitor; 5µl of DEPC-treated H_2O; 1µl of ImProm-II™ Reverse Trascriptase. The solution was always mixed after adding each component. The reaction was incubated on a thermocycler (MJ research Inc., Quebec, Canada) at 25°C for 5 minutes, 42°C for 1 hour, followed by final step of inactivation of reverse transcriptase at 70°C for 15 minutes. The product was verified on electrophoresis gel.

3.6.5 Real-time quantitative PCR (qPCR)

Single strand cDNA was amplified with the qPCR method in order to quantify mRNA relative expression level of *Glut1*. Normally, the qPCR was placed in white propylene 96 well plate and performed in LightCycler® 480 machine (Roche, Penzberg, Germany). 10µl volume reactions were prepared with a final concentrations of 1X Fermentas Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific, Waltham, USA), 0.15 mM of each primer and 1 µl cDNA, which dilution from the reverse transcription product (3.6.4) was adjusted according to the efficiency of the primers used. The efficiency, reproducibility of the primers used for the assay and the dynamic range of SYBR Green assay was determined by constructing a standard curve based on serial dilutions of cDNA. The standard curve was constructed by LightCycler® 480 Software 1.5 plotting the dilution factor against the threshold cycle (C_T) value obtained during amplification of each sample. qPCR assays conditions were chosen when the standard curve showed efficiency value (E)

of approximately 100%. After the optimization, qPCR was performed for each sample in three technical replicates and a minimum of three biological replicates. The relative expression of each sample was analysed as $2^{\Delta CT}$ where $\Delta C_T = C_T$ (calibrator) – C_T (test) and statistical comparison of the data was done performing an ANOVA analysis and post hoc tests (GraphPad Software, Inc., La Jolla, California, USA).

3.6.5.1 Glut1 relative quantification by qPCR

QPCR was used to quantify the mRNA expression levels of *Glut1* in several fly stocks used in our study. Changes in expression were calculated by relative quantification, where expression level of *Glut1* was normalized for expression level of ribosomal protein L32 (*RpL32*), ubiquitously expressed in the fruit fly. The primers for *Glut1* were designed using IDT® SciTools RealTime PCR software (http://eu.idtdna.com/scitools/Applications/RealTimePCR/; Integrated DNA Technologies, Inc., Coralville, Iowa, USA), the primers for *RpL32* (Table 26). *Glut1* and *RpL32* qPCRs, prepared as separated mix as reported in 3.6.5, were performed simultaneously with LightCycler® 480 machine following the protocol in Table 6.

Table 6. qPCR protocol for *Glut1* and *RpL32*.

	Analysis mode	Target (°C)	Acquisition mode	Hold	Ramp rate (°C/s)	Acquisition (per °C)
Pre-incubation (1 x)	−	95	none	10‘	4.4	−
Amplification (38 x)	quantification	95	none	15‘‘	4.4	−
		62	none	30‘‘	2.2	−
		72	single	30‘‘	4.4	−
Melting curve (1 x)	melting curve	95	none	5‘‘	4.4	−
		65	none	30‘‘	2.2	−
		97	continuous	−	−	5°C
Cooling (1 x)	−	40	none	30‘‘	2.2	−

x *refers to cycle.* ‘ *refers to minute;* ‘‘ *refers to second.*

Data analysis was done as reported in 3.6.5.

3.6.6 Lifespan assay

The *Drosophila* lifespan was assessed approximately on 100 female flies for each fly strain analysed. The flies were collected within 24 hours of emergence and 10 females were allocated in a vial at 25°C. According to the genotype of interest, the flies were transferred daily or every 3-4 days in a new tube and the dead flies were counted. Survival curves were generated, data were analysed by using Kaplan–Meier method and statistical significance was tested by using log rank statistics software (GraphPad Software, Inc., La Jolla, CA, USA).

3.6.7 Pseudo-pupil Assay

The pseudo-pupil assay was assessed on flies expressing the genotype of interest at day 1 and day 7 post-eclosion. The tested fly was anesthetized, decapitated and the head affixed to a microscope slide with a drop of nail varnish. The pseudo pupil analysis was performed using either a Nikon Optiphot-2 (Nikon, Shinjuku, Japan) or an Olympus BH2 (Olympus, Tokyo, Japan) light microscopes with oil immersion optics and a 500X magnification. The number of rhabdomeres was counted for 20-30 ommatidia for each eye of 10-15 flies. Statistical comparison of the data was done performing an ANOVA analysis and post hoc tests (Statistica v 5.0, StatSoft, Inc., Tulsa, OU, USA).

3.6.8 Eclosion Assay

The eclosion assay was performed by setting up 10 independent crosses between 5 males carrying the *elav-GAL4* driver and 5 virgin females homozygous for the *UAS*-trasgene. Flies were left to mate for 5 days after which the parental flies were removed from the vial. The crosses were cleared each day and the numbers of males and females were counted until all of the viable F1 progeny had eclosed. The adult emergence percentage was calculated as a ratio between female and all progeny, where the females express the transgene of interest. The data collected were analysed performing an ANOVA (Statistica v 5.0, StatSoft, Inc., Tulsa, OU, USA).

3.7 Miscellaneous

3.7.1 Suppliers

The following companies were suppliers of laboratory consumables for this thesis:

- Bioline, London, UK (agarose gel electrophoresis reagents)
- Biorad, Hercules, CA, USA (WB kit and ladders)
- Bloomington Drosophila Stock Center, Indiana University, Bloomington, IN, USA (fly strains)
- Eppendorf, Hambuerg, Germany (General)
- KAPA Biosystem, Woburn, MA, USA (buffer A and Taq polymerase)
- Life Technologies, Carlsbad, CA, USA (PCR reagent and machine; Capillary electrophoresis equipment and reagents; RNA extraction and DNase treatment; cell media)
- Promega, Madison, WI, USA (dNTPs; ImProm-II™ Reverse Transcriptase kit)
- Sigma-Aldrich, St Louis, MO, USA (General)

3.7.2 Buffers

All pH values at 25°C.

- 10X LD PCR buffer=50mM Tris-HCl (pH 8.8), 12.5mM ammonium sulphate, 1.4mM magnesium chloride, 125µg/ml BSA, 7.5mM 2-mercaptoethanol, 200µM of each dNTP
- 5X TBE (pH 8.3): Tris Base 0.45M, Boric acid 0.4M, EDTA 0.1M
- LB: 20 mM/L Tris-acetate, 1 mM/L EGTA, 1% Triton X-100, 10 mM/L β-glycerol phosphate, 5 mM/L sodium orthovandate, 10 mM/L β-glycerophosphate, 1 mM/L dithiothreitol, 50 ul Complete Mini Protease Inhibitor Cocktail (Roche)
- 4X WB loading buffer 4x: 62.5 mM Tris-HCl ph 6.8; 2% SDS; 0.1% Bromophenol-Blue; 10% Glycerol ; 10% β-mercaptoethanol

- 4X Stacking buffer: 0.5M Tris HCl pH 6.8; 0.4%v/v SDS
- 4X Separation buffer: 1.5M Tris HCl pH 8.8; 0.2% v/v SDS
- 5X Migration buffer: 250mM Tris; 1.9M Glycine; 0.5% v/v SDS
- 1X Transfer buffer: 20mM Tris; 150mM Glycine; 12% v/v methanol; 0.1% v/v SDS
- 10X Tris-Buffered Saline (TBS): 200mM; 1.5M NaCl; pH 7.6
- 1X TBS-Tween (TBS-T): 1X TBS; 0.1% v/v Tween 20
- Blocking buffer: 1X TBS-T, 5% w/v milk powder
- TBS-T with 1%w/v milk: 1X TBS-T; 0.1% w/v milk powder

CHAPTER FOUR: RESULTS

4.1 Characterisation of the HD cohorts

From now on we refer to the first HD sample cohort that we analysed in our study as HD cohort 1 and the other, which was provided in at second time, as HD cohort 2. The information provided for each cohort (given after completion of our analysis) included sex, the *HTT* gene CAG repeat length, AO of HD and main symptom(s) present at the disease onset according to the rater. The CAG mutation length was analysed by the local laboratory, where the sample was originally collected, and reanalysed by BioRep (BioRep, Milano); though this information was not provided for all samples. For our study the CAG mutation length of the *HTT* larger allele estimated by BioRep (BioRep, Milano) was considered and when this information was not available, the data from the local laboratory was used.

The clinical data for 495 samples of the HD cohort 1 were provided by EHDN after the completion of our analysis, except for 5 samples that were therefore excluded from our study. In the HD cohort 1 the CAG mutation length considered was in most of the cases provided by BioRep, except for 90 samples for which the data from the collecting laboratory were used. One sample carrying a 35 CAG repeat *HTT* allele, which by definition is not a disease allele, was included in the analysis due to ambiguity with the data provided by the collecting laboratory, where a 40 CAG repeat length mutation was identified. In the HD cohort 1 the expanded trinucleotide repeats ranged from 36 to 67 with a mean (±SD) of 44±4 CAGs, and AO ranged from 10 to 79 years, with a mean onset of 43±11 (SD) years. Two-hundred and fifty eight patients were diagnosed by motor disturbances (mean±SD motor AO= 44±11 years), 114 with psychiatric disturbances (mean±SD psychiatric AO= 39±10 years), 44 with cognitive decline (mean±SD cognitive AO= 41±12 years), 5 with symptoms such as weight loss or insomnia, 1 with oculomotor deficits and the remaining 73 with mixed symptoms (mean±SD mixed symptoms AO= 44±10 years) (Table 7). The HD cohort 1 included 263 men and 232 women.

Clinical data for 493 samples of the HD cohort 2 were provided by EHDN at the time of our analysis. The CAG mutation length considered was provided by BioRep (BioRep, Milano) for 425

samples and by the collecting laboratory for 68 samples. One sample was excluded from the analysis because was genotyped with only 29 CAG repeats units, which is not considered to cause HD by definition. In the HD cohort 2 the expanded trinucleotide repeats ranged from 37 to 77 with a mean (±SD) of 45±4 CAGs, and AO ranged from 6 to 83 years, with a mean onset of 44±13 (SD) years. Three hundred and sixty-two samples were diagnosed by motor disturbances (mean±SD motor AO= 45±13 years), 55 with psychiatric disturbances (mean±SD psychiatric AO= 41±12 years), 35 with cognitive decline (mean±SD cognitive AO= 42±15 years), 3 with symptoms such as weight loss or insomnia, 1 with oculomotor deficits and the remaining 36 with mixed symptoms (mean±SD mixed symptoms AO= 44±4 years) (Table 7). The HD cohort 2 included 227 men and 265 women.

Table 7. Main symptom estimated at HD onset for the samples of our cohorts.

Main symptom at onset	Cohort 1	Cohort 2	Cohort 1+2
Motor	258	364	620
Cognitive	44	35	79
Psychiatric	114	55	169
Oculomotor	1	1	2
Weight loss, insomnia	5	3	8
Mixed	73	36	109
Total	495	493	987

The clinical data in regards of the HD cohort 1 and 2 did not include any information of the ethnic origin of the samples, however a list of the provenience of the samples according to the recruiting clinical centre was provided. The location of the recruiting centres all over Europe suggests that the two cohorts are representative of different ethnic groups, but more detailed information are needed in order to exclude any possible bias towards ethnic groups or demographic origin (Figure 22).

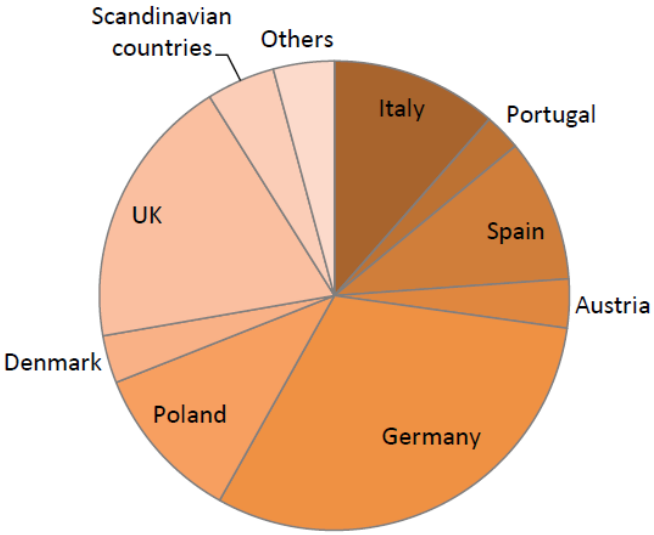


Figure 22. Location of the collecting centres of the REGISTRY HD samples in our study.

The percentage of each country is calculated on the number of the patients included in our study, recruited in a clinical centre in that country.

In order to asset the validity of our cohorts as well representative of the HD population we plotted the *HTT* CAG repeat length for the AO in HD cohort 1 and HD cohort 2 and it is clear the inverse correlation within the two variables, typical feature of HD (Figure 23).

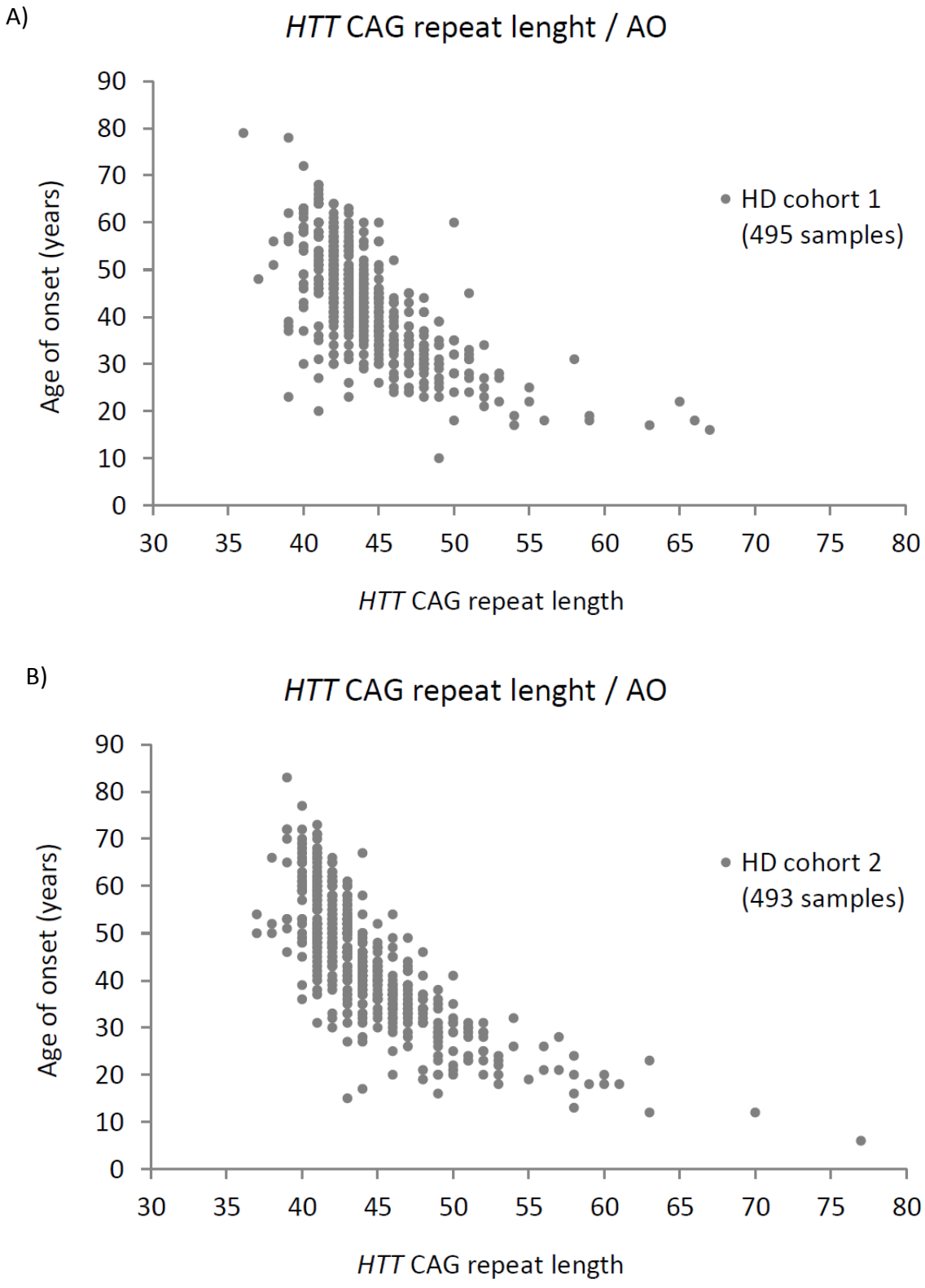


Figure 23 A, B. Scatter plots of AO and HTT CAG repeat length in HD cohort1 (A) and HD cohort 2 (B).

In order to investigate the size of the effect of *HTT* CAG repeat length on the variability of AO in HD cohort 1 and HD cohort 2 and in the two cohort together, we constructed a generalized linear model (GLM) using SPSS 20.0 (IBM Inc., Armonk, NY, USA) calculated using Type III statistics with Wald confidence intervals, with an identity link, assuming a normal distribution of the dependent variable. The AO was included in the model as Log2 transformed AO, which is assumed to follow a normal distribution (Figure 24) and better fits the correlation within the *HTT* CAG repeat length.

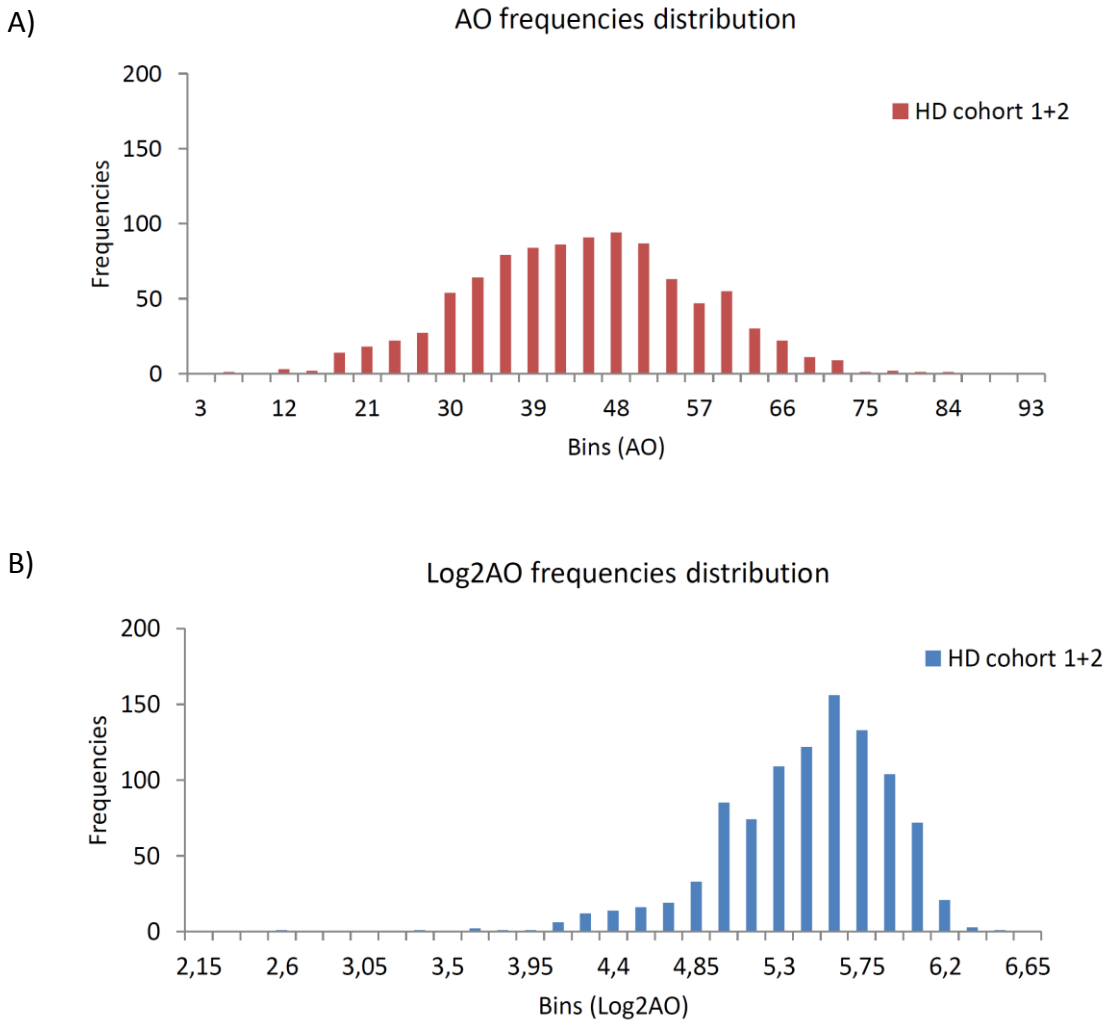


Figure 24 A, B. Histograms of age of onset (A) and Log2 transformed age of onset (B) frequencies distribution in our HD cohorts.

Log2 transformed AO of HD was the scalar dependent variable and *HTT* CAG repeats length the scalar predictor variable. As expected, the CAG repeat length is significantly contributing to the prediction of the AO in HD cohort 1, HD cohort 2 and HD cohorts together (Table 8).

Table 8. Effect of the HTT CAG repeat length on the age of onset in HD cohort 1 and 2.

	Mean (95%CI) (years per CAG repeat)	p-value
Effect for extra CAG repeat in HD cohort 1	−2.14 (−2.33 to −1.96)	$<5\times10^{-6}$
Effect for extra CAG repeat in HD cohort 2	−2.5 (−2.68 to −2.39)	$<5\times10^{-6}$
Effect for extra CAG repeat in HD cohort 1+2	-2.34 (−2.46 to −2.22)	$<5\times10^{-6}$

The inclusion into the GLM of the sex, as categorical predictor factor, is not associated to any significant effect in HD cohort 1 (p value = 0.332), HD cohort 2 (p value = 0.754) and HD cohort 1 and 2 together (p value = 0.392). Furthermore, we performed a one-way ANOVA to test if there is any correlation between major symptom(s) at onset and the *HTT* CAG repeat length and no significant association was found in HD cohort 1 (p value = 0.828), HD cohort 2 (p value = 0.569) and HD cohort 1 and 2 (p value = 0.851).

The two HD cohorts independently and taken together are well representative of the variability of AO in HD, which is main attributable to the *HTT* CAG repeat length, within an inverse correlation between the two variables. Furthermore, the two HD cohorts independently and taken together show no bias towards the mayor symptom(s) at onset or sex.

4.2 β-defensin CNV distribution in HD cohort 1

We analysed β-defensin CNV in 500 samples from HD cohort 1. The β-defensin copy number for each sample was inferred from six values, resulting from the β-defensin assay, which were corrected for inter-experiment variability using known controls (Aldhous *et al*. 2010) (Table

25). The quality of copy number calling was evaluated qualitatively by analysing histograms of the mean of the normalized ratio for each PRT (PRT107A, HSPD21) expecting to have clusters around every integer value which indentified each copy number of the β-defensin locus (Figure 25, Figure 26). The histograms in Figure 25 and Figure 26 show clear clustering around integer values from each PRT. In order to adjust the copy number calling for contiguous values, the β-defensin method combines the two PRTs with the 5DELR assay, which allows distinguishing between contiguous copy number calls producing different value for even and odd copy number samples. The reproducibility of the β-defensin assay was tested comparing the raw data from the control samples, which were used in each reaction. The control samples showed similar raw PRT ratios from separate reactions confirming the reproducibility of the β-defensin assay; the variance of their raw values could be explained by procedural variability (sample dilution batch, PCR mix) (Figure 27).

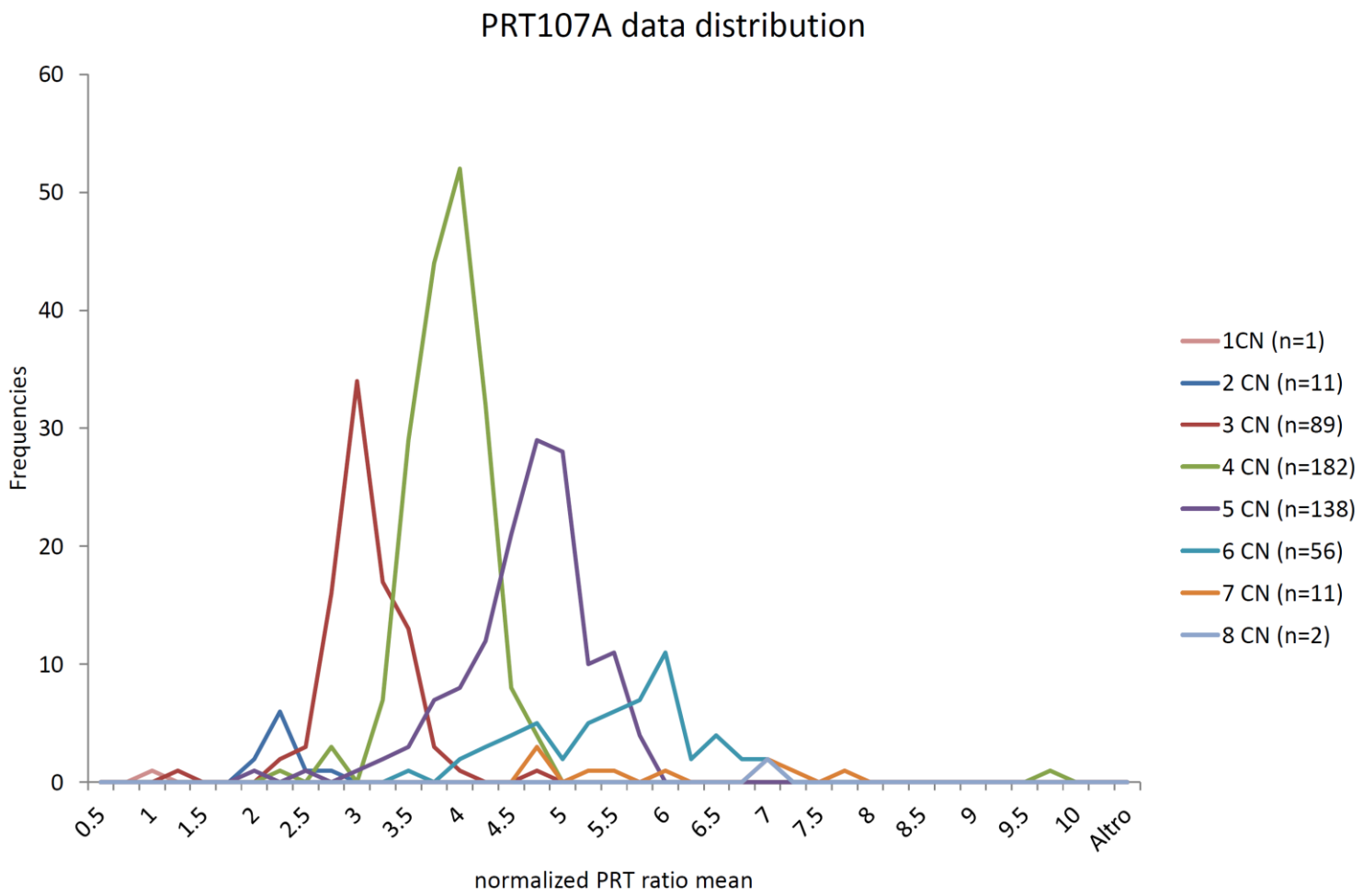


Figure 25. Histogram of PRT107A ratios frequencies plotted for final β-defensin copy number in HD cohort 1.

"CN" refers to final β-defensin copy number, as determined by the maximum likelihood associated to each PRT raw ratio. "n" to the number of individuals.

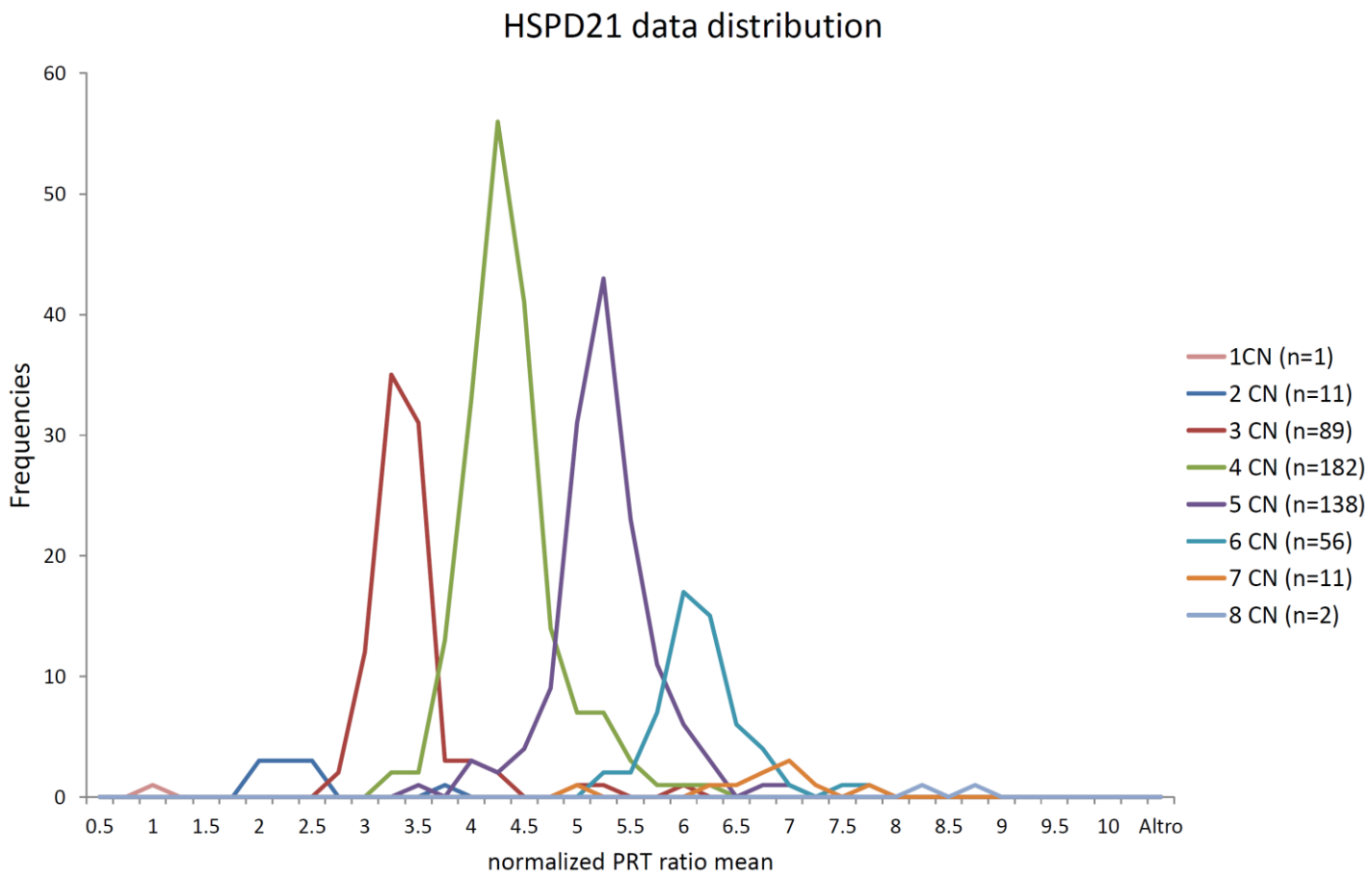
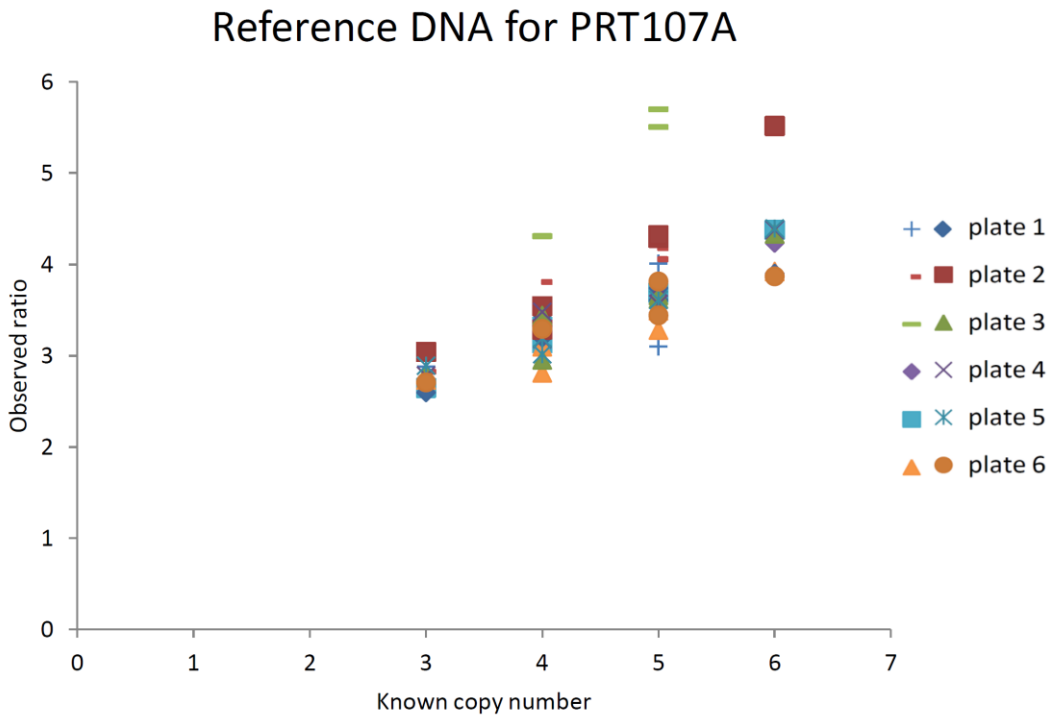


Figure 26. Histogram of HSPD21 ratios frequencies plotted for final β-defensin copy number in HD cohort 1.

"CN" refers to final β-defensin copy number, as determined by the maximum likelihood associated to each PRT raw ratio. "n" to the number of individuals.

A)



B)

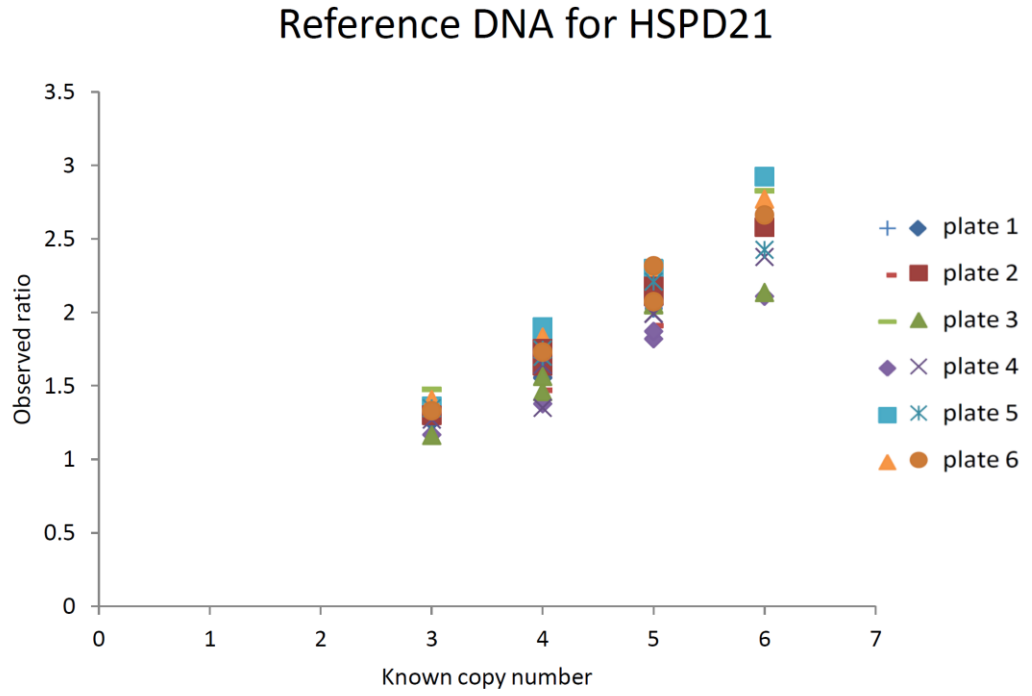


Figure 27 A, B. Scatter plot of raw ratios from 72 controls samples used in each PRT107A (A) and HSPD21 (B) reaction.

It represents repeat testing of six DNA samples, available from the European Collection of Cell Cultures, CO088 4 copies, CO207 5 copies, CO849 6 copies, CO913 3 copies, CO940 4 copies, CO969 5 copies. The different symbols for each plate represent the PRT replicates done with different labelled primers.

For all the samples of HD cohort 1 we calculated p value that reflects the confidence in the inferred copy number compared to all the other copy number calls using the maximum likelihood approach. We determined β-defensin copy number with p value < 0.05 for 495 and among these 99% have copy number call with p value < 0.01. Five samples were not included in the analysis because having a copy number with p value > 0.05.

In order to investigate the effect of β-defensin CNV on the AO in HD, we excluded 5 samples because no clinical data were available at the time of the analysis. Firstly, we analysed the β-defensin copy number distribution in the HD cohort 1 dividing the cohort into three groups according to the age of onset: juvenile, AO ≤ 20 years; typical, AO between 20 and 50 years; late, AO ≥ 51 years (Wexler *et al*. 2004) (Table 9).

Table 9. β-defensin CNV genotypes in each AO class, in the HD cohort, and the European population.

β-defensin copy number	Juvenile AO HD group	Typical AO HD group	Late AO HD group	HD cohort	European population
1	0	1	0	1	0
2	0	8	3	11	8
3	3	62	24	89	70
4	2	129	51	182	205
5	5	99	34	138	131
6	1	44	11	56	44
7	0	8	3	11	9
8	0	1	1	2	4
9	0	0	0	0	1
n	11	352	127	490	472

We carried out a one-way ANOVA to test whether there was any variation in β-defensin CNV distribution among each group, and all the groups and the European population without HD, analysed in a previous study (Fode *et al*. 2011). The distribution of no AO class had a β-defensin CNV distribution that was significantly different from any other class or the European non-HD population and, the β-defensin CNV distribution in the whole HD cohort 1 was not significantly different from the European population without HD (one-way ANOVA, p value = 0.91) (Figure 28) (Fode *et al*. 2011).

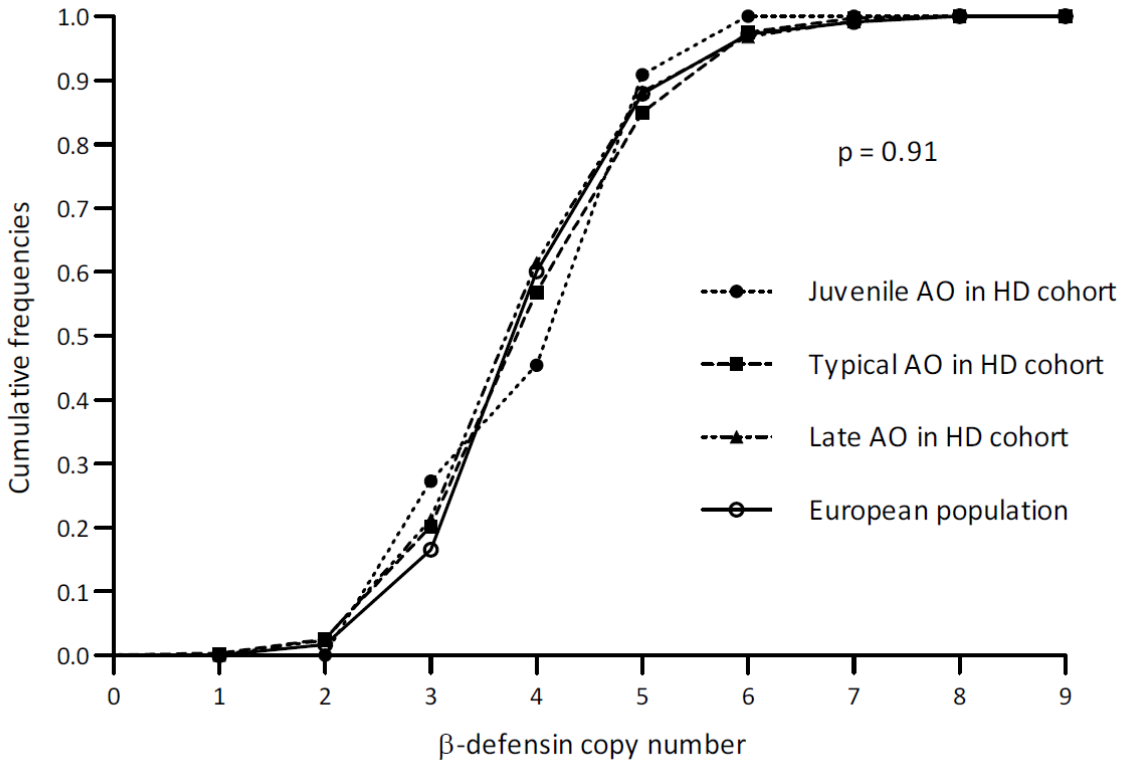


Figure 28. Cumulative frequencies of β-defensin CNV genotypes in HD individuals subsets and European population.

HD subjects were grouped according to the AO in juvenile AO, typical AO, late AO and of the non-HD European population. The p-value is two-sided, calculated by one-way ANOVA.

We constructed a GLM to explore the effect of CAG repeat length and β-defensin copy number, as ordinal predictor variable, on AO in HD. The addition of CAG repeat length into the model improved the prediction of AO (p value < 5×10−6, Table 4), but subsequent incorporation of β-defensin copy number into the model found no significant improvement in the prediction of AO (p value = 0.41, Table 10). No significant effect was found adding the gender to the model (p value = 0.34). We also used a GLM to investigate the effect of β-defensin copy number and CAG repeat length on the major estimated symptom(s) at onset of HD (categorical dependent variable) and no significant effect with any of the symptoms at the onset and CAG repeat length (p value = 0.76) nor β-defensin copy number (p value = 0.85) was found. Thus, it appeared that β-defensin CNV does not significantly affect the AO of HD in HD cohort 1.

Table 10. Effect of the β-defensin CNV on the age of onset in HD.

	Mean (95%CI) (years)	p-value
Effect per extra CAG repeat	−2.17 (−2.36 to −1.98)	$<5\times10^{-6}$
Effect per extra copy of *DEFB4*	−0.28 (−0.96 to 0.4)	0.415

4.3 *SLC2A3* CNV in HD

4.3.1 *SLC2A3* CNV distribution in HD cohort 1 and 2

SLC2A3 assay was performed and analysed separately in HD cohort 1, HD cohort 2 and afterwards analysed in both cohorts together. *SLC2A3* copy number was defined by a mean ratio corrected for inter-experimental variability using known controls (Table 25), whose *SLC2A3* copy number was determined in a previous study using the same PRT assay and recombination assays (Veal *et al.* 2013). The reproducibility of the *SLC2A3* assay was tested comparing the raw data from the control samples (Table 25), which were used in each reaction. The control samples show similar raw PRT ratios from separate reactions confirming the reproducibility of the *SLC2A3* assay (Figure 29).

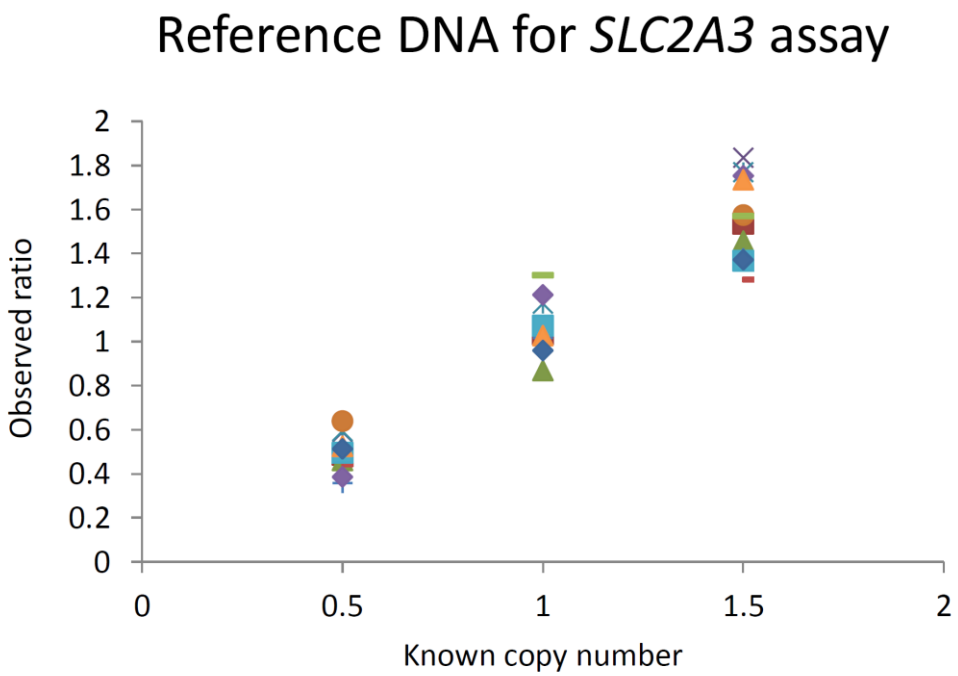
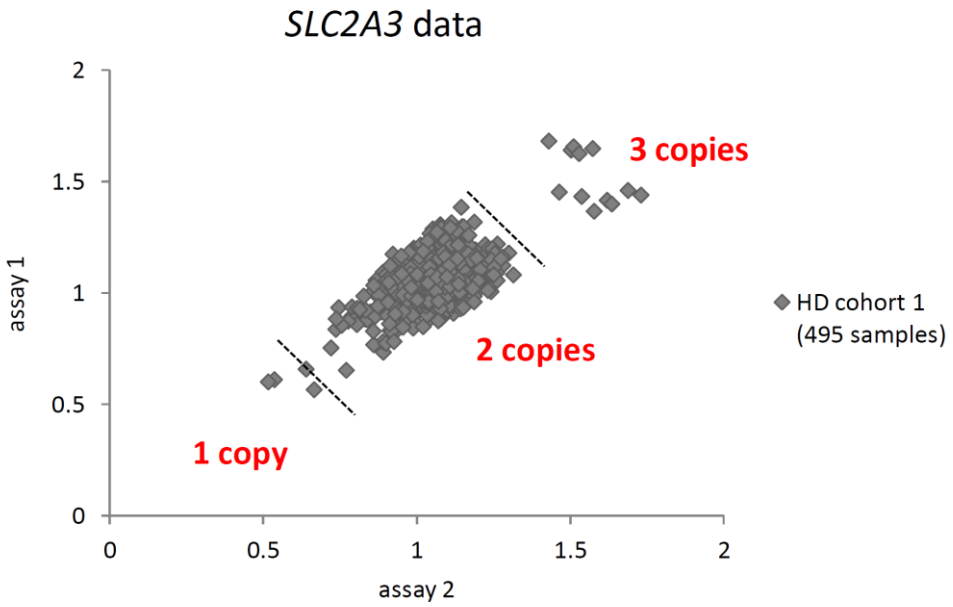


Figure 29. Scatter plot of raw ratios from 36 controls samples used in each *SLC2A3* assay.

It represents repeat testing of 3 DNA samples, NA19920 3 copies (from HapMap Panel), NA11840 2 copies (from CEPH panel), L31 1 copy (from Leicester panel). The different symbols represent different SLC2A3 assay reactions.

The copy number calling was evaluated qualitatively by plotting the normalized ratio for each replicate of the *SLC2A3* assay expecting to have clusters around every integer value which indentified each copy number of the *SLC2A3* locus. Two clear clusters were identified after plotting values from technical replicates from HD cohort 1 genotyping, allowing defining technical cut-off for copy number calling. The first threshold used to distinguish between 1 copy and 2 copies was set at 0.65: any sample with a mean normalized ratio equal or less than 0.65 was inferred to be 1 copy number sample (Figure 30-A). The second threshold was set up to distinguish between 2 and 3 copies: any sample with a normalized ratio equal or higher than 1.35 was called 3 copies (Figure 30-A). Samples with ratio between 0.65 and 1.35 were called as 2 copy number samples (Figure 30-A). Plotting the data from two technical replicates of *SLC2A3* assay performed on HD cohort 2 gave the same distribution of normalized values. Therefore, the cut-offs used to infer *SLC2A3* copy number in HD cohort 1 were identical in HD cohort 2 (Figure 30-B). Eventually, the same cut-offs were confirmed combining the HD cohort 1 and HD cohort 2 normalized value plots together.

A)



B)

Figure 30 A-B. SLC2A3 data plots in HD cohort 1 (A), HD cohort 2(B).

For each cohort normalized ratios from assay replicates were plotted together. Clear clusters allowed to identify technical cut off for inferring integer SLC2A3 copy number. The dot lines represent the technical cut-offs used to distinguished between 1 and 2 SLC2A3 copy number and 2 and 3 SLC2A3 copies. The sample in B) showing a ratio of ~ 2 was considered an outlier.

From HD cohort 1, we successfully genotyped 495 samples and we found 4 deletion carriers, 12 duplication carriers and 479 individuals with normal diploid set. Five samples were excluded from the study because repeated analysis gave results that were rejected according our experimental condition. Firstly we compared *SLC2A3* CNV distribution in HD cohort 1 and the British population (Reekie 2011) and no significant difference was found between the two groups (t-test two-tailed, p value = 0.38) (Table 12). As previously described, we constructed a GLM to investigate the effect of CAG repeat length and *SLC2A3* copy number (as ordinal predictor variable) on AO in HD. As expected, including the CAG length into the model bettered the prediction of AO (p value < 5×10^{-6}, Table 11) and the subsequent incorporation of *SLC2A3* copy number into the model was associated to a significant improvement in the prediction of AO (p value = 0.038, Table 11). No significant effect was found adding the sex to the model (p value = 0.30). Furthermore, we used a GLM to investigate the effect of *SLC2A3* copy number and CAG repeat length on the major estimated symptom(s) at the onset of HD and no significant effect with any of the symptoms at the onset diagnosis and CAG repeat length (p value = 0.69) nor *SLC2A3* copy number (p value = 0.49) was found.

Table 11. Effect of the *SLC2A3* CNV on the age of onset in HD cohort 1

	Mean (95%CI) (years)	p value
Effect per extra CAG repeat	-2,20 (-2,39 to -2,04)	$<5\times10^{-6}$
Effect per extra copy of *SLC2A3*	4.53 (0.23 to 9.23)	0.038

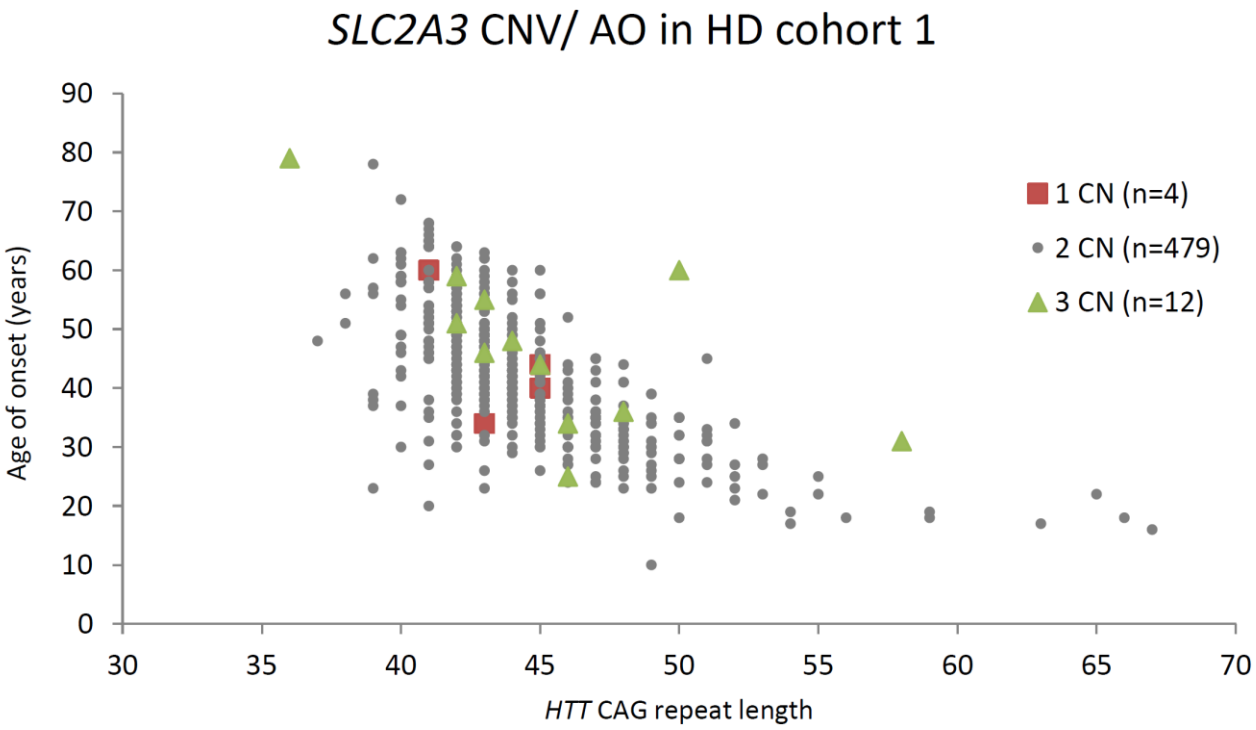


Figure 31. Scatter plot of AO and *HTT* CAG repeat length in our HD cohort 1 within *SLC2A3* copy number genotype.

Grey dots indicate individuals with 2 copies, green triangles with 3 copies and red squares with 1 copy of SLC2A3.

In order to investigate the effect of *SLC2A3* CNV on the AO in HD we analysed *SLC2A3* CNV distribution in additional 500 samples of the HD cohort 2. We determined *SLC2A3* copy number in 499 samples of HD cohort 2 using and we identified 1 deletion, 24 duplications, with the remaining 474 samples having a *SLC2A3* diploid set. One sample was excluded from our analysis because it showed a copy number of 4, which was not verified due to the absence of any standard control of copy number higher than 3 in our assay or of other methods to test the origin and structure of this CNV event. Firstly we tested *SLC2A3* CNV distribution in HD cohort 2 compared to HD cohort 1 and the British population. There was no significant difference between HD cohort 2 and HD cohort 1 (t-test two-tailed, p value = 0.99) nor the British population (t-test two-tailed, p value = 0.38), moreover the two cohorts taken together were not significantly different from the British population (t-test two-tailed, p value= 0.42) (Table 12).

Table 12. SLC2A3 CNV genotypes in the HD cohorts and the British population.

SLC2A3 copy number	HD Cohort 1		HD Cohort 2		HD Cohorts 1+2		British population	
1	4	(0.8%)	1	(0.2%)	5	(0.5%)	58	(0.9%)
2	479	(96.8%)	474	(95%)	953	(95.8%)	6016	(95.4%)
3	12	(2.4%)	24	(4.8%)	37	(3.7%)	234	(3.7%)

To measure the effect of the *SLC2A3* CNV distribution on AO in HD cohort 2, 7 samples (with *SLC2A3* copy number of 2) were excluded from the study due to the lack of clinical data at the moment of the analysis. As previously described (2.4), we constructed a GLM to investigate the effect of *HTT* CAG repeat length and *SLC2A3* copy number (as ordinal predictor variable) on AO of HD in HD cohort 2. Including the *HTT* CAG repeat length into the model bettered the prediction of AO (p value < 5×10^{-6}, Table 13). The subsequent incorporation of *SLC2A3* copy number into the model did not give a significant improvement in the prediction of AO of HD (p value = 0.325, Table 13). No significant effect was found including the sex into the model (p value = 0.768). Furthermore, we used a GLM to investigate the effect of *SLC2A3* copy number and CAG repeat length on the major estimated symptom(s) at the onset of HD and no significant effect with any of the symptoms at the onset diagnosis and CAG repeat length (p value = 0.66) nor *SLC2A3* copy number (p value = 0.22) was found.

Table 13. Effect of the *SLC2A3* CNV on the age of onset in HD cohort 2.

	Mean (95%CI) (years)	p value
Effect per extra CAG repeat	−2.47 (−2.63 to −2.33)	$<5\times10^{-6}$
Effect per extra copy of *SLC2A3*	1.6 (−1.53 to 4.99)	0.325

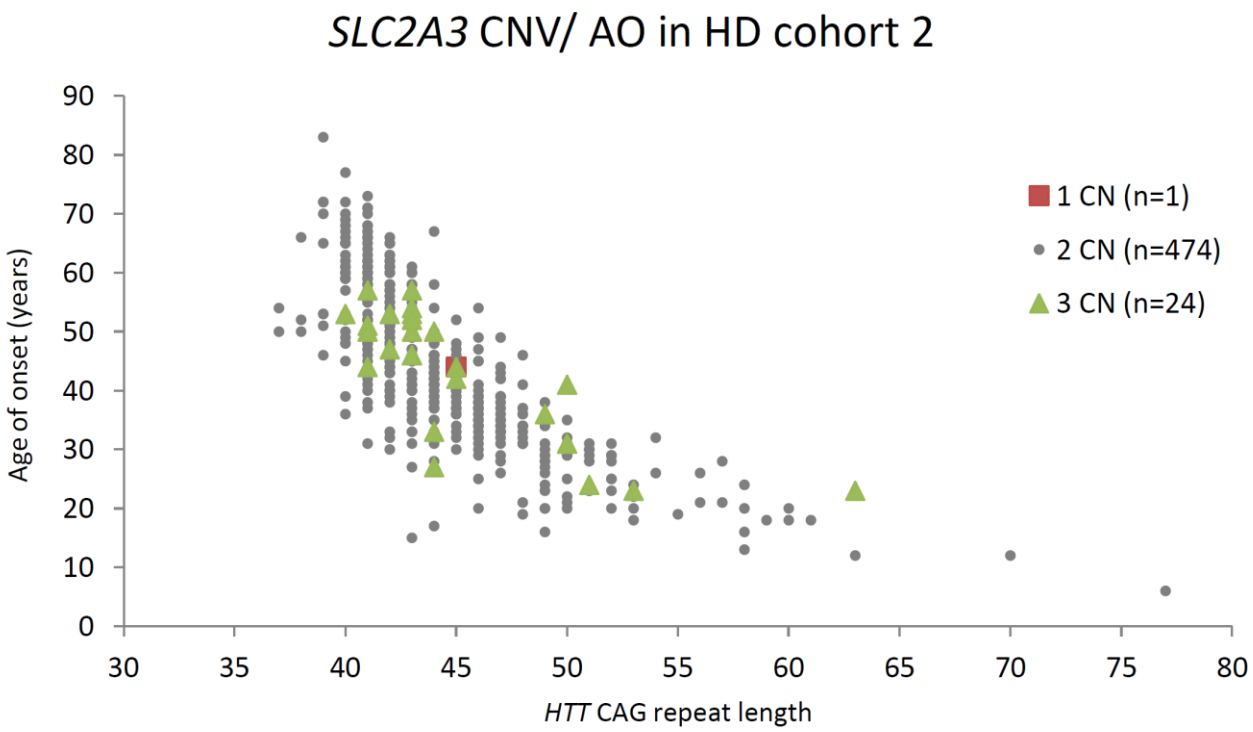


Figure 32. Scatter plot of AO and *HTT* CAG repeat length in our HD cohort 2 within *SLC2A3* copy number genotype.

Grey dots indicate individuals with 2 copies, green triangles with 3 copies and red squares with 1 copy of SLC2A3.

The negative correlation between *SLC2A3* CNV distribution in HD cohort 2 and the positive correlation in HD cohort 1 suggested that the two cohorts, formed in order to be well-representative samples of the HD population (considering the *HTT* CAG repeat length and also AO of HD), did not behave as a good estimator for studying *SLC2A3* CNV as a genetic modifier of AO in HD. Furthermore, it was important to consider the size of the genetic variation analysed: copy number change of *SLC2A3*, namely deletion or duplication, was present in the HD cohort 1 and 2 taken together (996 individuals) and in the British population (6308 individuals) as a rare event involving only on average ~4.5% of the samples in each of these groups. The size of this genetic variation, which effect on AO of HD was analysed as covariant with the *HTT* CAG mutation length, could negatively affect the power of the analysis in the single HD cohort 1 or HD cohort 2, probably leading to false positive or false negative associations. Therefore, considering that the two cohorts were not distinct population but samples of the same population, it was appropriate

to combine them together in order to have a larger sample which could improve the power of our analysis to describe the effect of *SLC2A3* CNV on AO of HD. Performing a GLM to studying the effect of *SLC2A3* CNV in the HD cohort 1 and 2 taken together (Figure 33), we found that *SLC2A3* CNV was associated to a significant improvement in the prediction of AO, higher than in regards of the HD cohort 1 alone (p value = 0.028, Table 14).

Table 14. Effect of the *SLC2A3* CNV on the age of onset in HD cohort 1 and 2.

	Mean (95%CI) (years)	p value
Effect per extra CAG repeat	−2.37 (−2.48 to −2.24)	$<5\times10^{-6}$
Effect per extra copy of *SLC2A3*	2.89 (0.28 to 5.61)	0.028

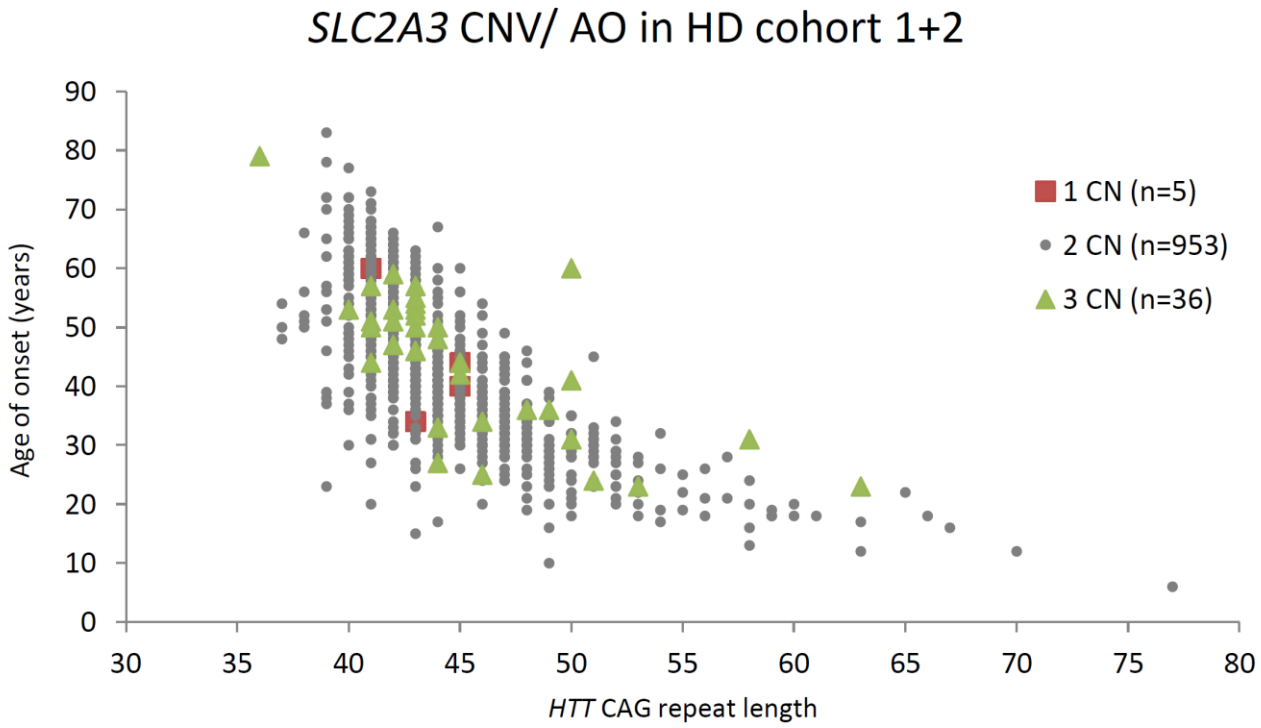


Figure 33. Scatter plot of AO and HTT CAG repeat length in our HD cohort 1 and 2 within *SLC2A3* copy number genotype.

Grey dots indicate individuals with 2 copies, green triangles with 3 copies and red squares with one copy of SLC2A3.

No significant effect was found adding to the GLM the gender (p value = 0.39). Constructing a GLM for the main symptom(s) at onset in HD cohort 1 and 2 together, no significant effect for any of the symptoms detected at onset of HD was found incorporating CAG repeat length (p value = 0.65) nor *SLC2A3* CNV (p value = 0.074) to the model. Thus, it appears that *SLC2A3* CNV can affect the AO in HD where individuals bearing 3 *SLC2A3* copy number showed a delay of the AO up to ~ 6 years compared to individuals with 1 or 2 *SLC2A3* copy number, with no bias towards sex or major symptom at onset.

4.3.2 Analysis of *SLC2A3* CNV in patient cell lines

In order to understand the significant effect of *SLC2A3* CNV on the AO in our HD cohorts, we investigated the role of CNV on *SLC2A3* gene expression quantifying GLUT3 (encoded by *SLC2A3*) expression level. For this purpose we collected 15 LCLs originated from different samples of our two cohorts, 5 for each *SLC2A3* copy number. Among the individuals carrying 1 copy of *SLC2A3* the mean AO was 44.4 years (SD± 9.6) and the mean CAG mutation length was 43.8 (SD± 1.8); among the individuals carrying 2 copy of *SLC2A3* the average AO was 47 years (SD± 8.2) and the mean CAG mutation length was 42.6 (SD± 1.9); among the individuals carrying 3 copy of *SLC2A3* the average AO was 45 years (SD± 10.5) and the mean CAG mutation length was 43.8 (SD± 0.8) (Table 15).

Table 15. Clinical data and *SLC2A3* CNV genotype related to LCLs.

LCLs code	Age of onset (Mean ±SD)	CAG mutation length (Mean ±SD)	*SLC2A3* copy number
1	40	45	1
2	34	43	1
4	60	41	1
8	44	45	1
11	44	45	1
	(44.4 ± 9.6)	(43.8 ± 1.8)	
6	34	46	2
8	51	42	2
10	48	42	2
13	48	42	2
14	56	41	2
	(47 ± 8.2)	(42.6 ± 1.9)	
3	44	45	3
5	50	44	3
9	53	43	3
12	27	44	3
15	50	43	3
	(45 ± 10.5)	(43.8± 0.8)	

(Mean ± SD) refers to mean and standard deviation values (as indicated) for each SLC2A3 copy number group.

4.3.2.1 Immuno-quantification of GLUT3 protein level in LCLs

We quantified GLUT3 expression levels in our LCLs by immunoblotting analysis. According to a gene dosage effect we expected to detect GLUT3 protein level in direct correlation with *SLC2A3* copy number. One protein lysates was prepared for each cell line, after 72 hours incubation in normal media. Immunoblot experimental replicates were performed to reduce the effect of technical variability on the final analysis. We analysed the protein lysates of each sample in 5 different blots, except for LCL 3 and LCL 12, which were analysed in 4 different gels. The GLUT3 band was always detected at 45 kilo Dalton (kDa) and the α-Tubulin band, which was used to normalize the GLUT3 expression signal, at 55 kDa (Figure 34**,**Table 16).

Table 16. Descriptive data of GLUT3 relative protein level in the 3 CNV groups.

SLC2A3 copy number	Observations	Mean relative intensity	Std.dev.	Median	Min	Max
1	24	0.66	0.24	0.67	0.24	1.12
2	25	0.58	0.25	0.56	0.18	1.15
3	24	0.82	0.30	0.89	0.31	1.52

Std.dev indicates the standard deviation. Min refers to the minimum value. Max refers to the maximum value.

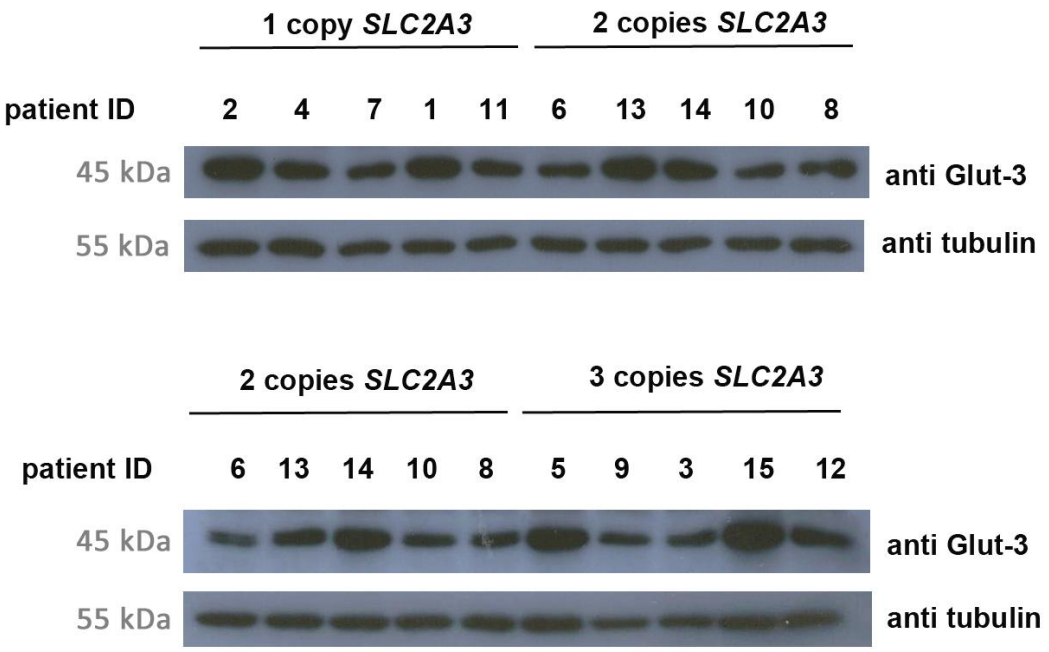


Figure 34. GLUT3 expression levels in LCLs by immunoblot.

Mixed effects linear regression was used to explore the association between the normalized intensity ratios, indicating the relative GLUT3 protein expression level, with *SLC2A3* copy number associated to each cell line. The analysis was designed to take into account the gel where the samples were loaded as a source of the experimental variability intra-replicates. *SLC2A3* copy number was found to be significantly associated with the protein level (p value = 0.02, Figure

35, Table 17). Furthermore we compared the relative GLUT3 expression levels among different *SLC2A3* CNV groups. The relative protein expression level for the group with *SLC2A3* CN=3 is significantly different to *SLC2A3* CN=1 (p value = 0.02, Table 18) (Figure 36) and *SLC2A3* CN=2 (p value <0.001, Figure 37) (Figure 37). No significant difference was found comparing the relative protein expression level in the group with SLC2A3 CN=1 and SLC2A3 CN=2 (p value = 0.30,Table 18-Table 19) (Figure 37).

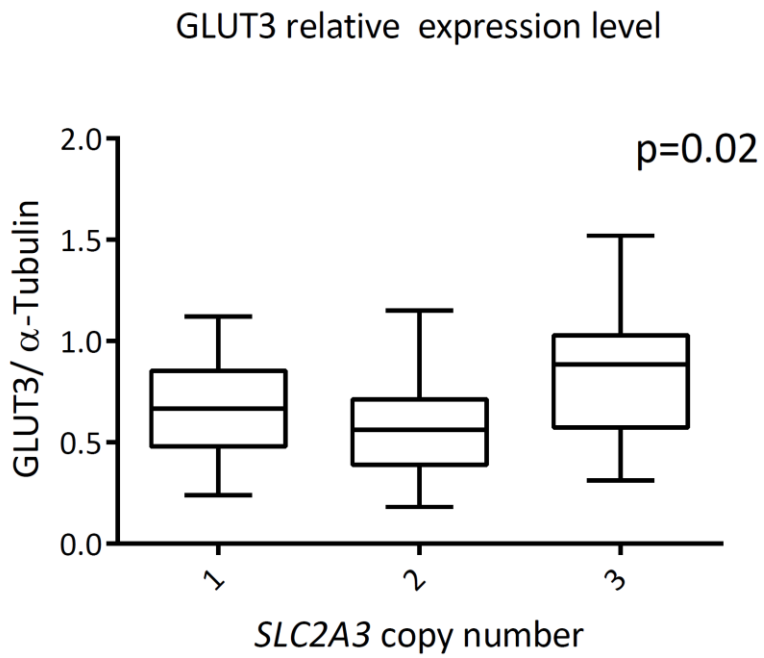


Figure 35. GLUT3 relative expression levels according to *SLC2A3* copy number.

The vertical bar indicates the minimum and the maximum value and the horizontal bar the mean value of each group. 1 SLC2A3 copy number, 3 SLC2A3 copy number: 24 observations each; 2 SLC2A3 copy number: 25 observations. Statistical comparison by mixed effects linear regression.

Table 17. Effect of *SLC2A3* copy number on protein expression level.

	Mean (95%CI) (GLUT3/ α-Tubulin intensity ratio)	p value
Effect per extra copy of *SLC2A3*	0.09 (0.02 to 0.17)	0.02

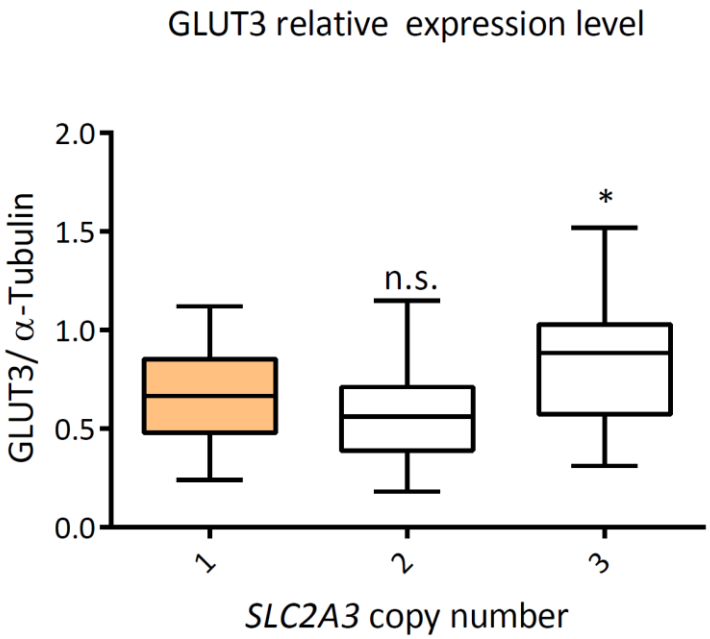


Figure 36. GLUT3 protein level in LCLs with 1 *SLC2A3* copy number compared to the other copy number classes.

The vertical bar indicates the minimum and the maximum value and the horizontal bar the mean value of each group. 1 SLC2A3 copy number, 3 SLC2A3 copy number: 24 observations each; 2 SLC2A3 copy number: 25 observations. Statistical comparison versus GLUT3 relative expression in 1 SLC2A3 copy number (orange box) by mixed effects linear regression.

Table 18. Effect of *SLC2A3* copy number 1 compared to the other CNV classes on protein expression level.

Effect compared to *SLC2A3* CN=1	Mean (95%CI) (GLUT3/ α-Tubulin intensity ratio)	p value
SLC2A3 CN=2	-0.07 (-0.22 to 0.07)	0.30
SLC2A3 CN=3	0.18 (0.04 to 0.33)	0.01

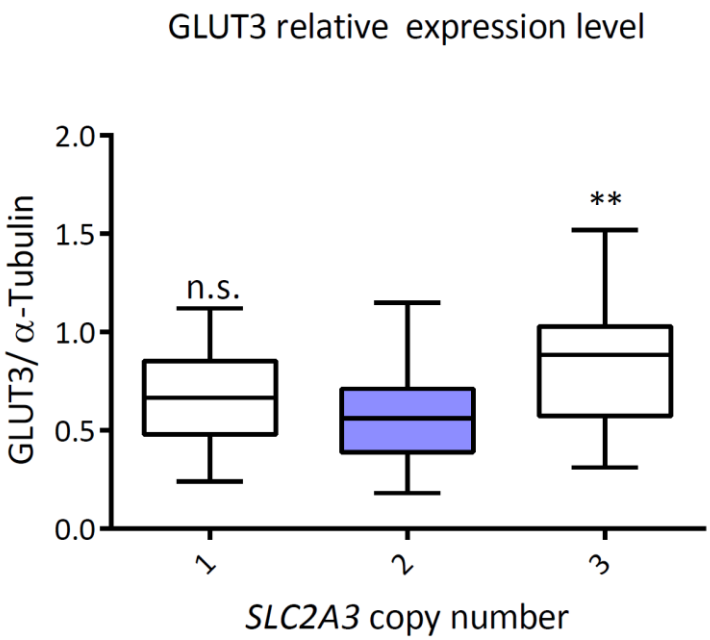


Figure 37. GLUT3 protein level in LCLs with 2 *SLC2A3* copy number compared to the other copy number classes.

The vertical bar indicates the minimum and the maximum value and the horizontal bar the mean value of each group. 24 observations each for 1 SLC2A3 copy number and 3 SLC2A3 copy number, 25 observations for 2 SLC2A3 copy number. Statistical comparison versus GLUT3 relative expression in 2 SLC2A3 copy number (blue box) by mixed effects linear regression.

Table 19. Effect of *SLC2A3* copy number 2 compared to the other CNV classes on protein expression level.

Effect compared to *SLC2A3* CN=2	Mean (95%CI) (GLUT3/ α-Tubulin intensity ratio)	p value
SLC2A3 CN=1	0.07 (-0.07 to 0.22)	0.30
SLC2A3 CN=3	0.26 (0.12 to 0.40)	<0.001

4.4 Functional study of the neuronal glucose transporter Glut1 in a Drosophila model of HD

Considering the significant correlation between *SLC2A3* CNV and the AO of HD in the tested cohorts and the significant effect of *SLC2A3* CNV on GLUT3 protein expression level in cell lines (LCLs) from HD individuals, we decided to study the function of the glucose transport in an *in vivo* model of HD. The study aimed to investigate if genetic alterations of a main neuronal glucose transporter could modulate the HD phenotype in the model and, specifically, if an increased expression of this transporter can have a protective effect against HD progression. For our purposes we decided to use a *Drosophila* HD model that within a central nervous system allows one to investigate the effect of selected gene expression specifically in the brain. Furthermore, the application of the *Drosophila* HD model includes the possibility to test the effect of a transgene through several metrics, such as survival, neurodegeneration in the eye and eclosion (emergence of the adult fly from the pupal case).

There are several sugar transporters annotated in the fruit fly genome (FlyBase, http://flybase.org/) and from those a number of glucose and sugar transporter amino acid sequences were selected in Uniprot, a protein database (http://www.uniprot.org/), and compared through multiple alignment with the main human GLUTs expressed in the brain using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). From the analysis of this multiple alignment (Figure 38), Drosophila Glut1 (Q6NNA9_DROME, Uniprot code) showed homology to human GLUT3 (GTR3_HUMAN, Uniprot code) (encoded by *SLC2A3*), sharing 46% of amino acid identity with its protein sequence (Figure 39).

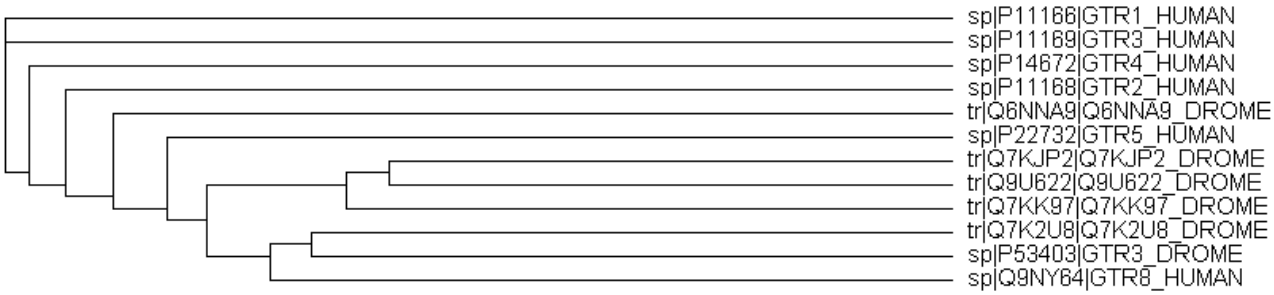


Figure 38. Dendogram of human GLUT1-5, 8 protein sequences and several sugar transporters in *Drosophila*.

The proteins are indicated with their Uniprot code.

```
1    MAFLCAPGLTFFLTYSIFSAVLGMLQFGYNTGVINAPEKNIENFMKDVYKDRYGEDISEE   60  Q6NNA9   Q6NNA9_DROME
1    ---MGTQKVTPALIFAITVATIGSFQFGYNTGVINAPEKIIKEFINKTLTDKGNAPPSEV   57  P11169   GTR3_HUMAN
        : :  :*  * ::*  *.:* :************** *::*::.. .*:     **

61   FIQQLYSVAVSIFAIGGMLGGFSGGWMANRFGRKGGLLLNNVLGIAGACLMGFTKVSHSY  120  Q6NNA9   Q6NNA9_DROME
58   LLTSLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAVTGGCFMGLCKVAKSV  117  P11169   GTR3_HUMAN
     :: .*:*::*:**::***:*.** * :.*****: .:*: *:*.::*.*:**: **::*

121  EMLFLGRFIIGVNCGLNTSLVPMYISEIAPLNLRGGLGTVNQLAVTVGLLLSQVLGIEQI  180  Q6NNA9   Q6NNA9_DROME
118  EMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVVGILVAQIFGLEFI  177  P11169   GTR3_HUMAN
     ***:***::**: *** *.:*****.**:*  ***.:**:***.::**:*::*::*:* *

181  LGTNEGWPILLGLAICPAILQLILLPVCPESPRYLLITKQWEEEARKALRRLRASGSVEE  240  Q6NNA9   Q6NNA9_DROME
178  LGSEELWPLLLGFTILPAILQSAALPFCPESPRFLLINRKEEENAKQILQRLWGTQDVSQ  237  P11169   GTR3_HUMAN
     **::* **:***::* *****   **.******:***.:: **:*:: *:** .: .*.:

241  DIEEMRAEERAQQSESHISTMELICSPTLRPPLIIGIVMQLSQQFSGINAVFYYSTSLFM  300  Q6NNA9   Q6NNA9_DROME
238  DIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFK  297  P11169   GTR3_HUMAN
     **:**: *.   ..*.:::.:**:   : * *:**.**:*****:***********.:*

301  SSGLTEESAKFATIGIGAIMVVMTLVSIPLMDRTGRRTLHLYGLGGMFIFSIFITISFLI  360  Q6NNA9   Q6NNA9_DROME
298  DAGVQE--PIYATIGAGVVNTIFTVVSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLL  355  P11169   GTR3_HUMAN
     .:*: *    :**** *.: .::*:**: *:::*:******: ***** : * ::*:*:*:

361  KEFFGYVQEMIDWMSYLSVVATLGFVVFFAVGPGSIPWMITAELFSQGPRPSAMAIAVLV  420  Q6NNA9   Q6NNA9_DROME
356  KDNY-------NGMSFVCIGAILVFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCS  408  P11169   GTR3_HUMAN
     *: :       : **::.: * * **.** :*** ***:*.**********:***:*

421  NWMANFVVGIGFPSMKTALENYTFLPFSVFLAIFWIFTYKKVPETKNKTFEEILALFRHN  480  Q6NNA9   Q6NNA9_DROME
409  NWTSNFLVGLLFPSAAHYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQ  468  P11169   GTR3_HUMAN
     ** :**:**: ***    *  *.*: *: **  *  **: *****: :***:*   *. :

481  NGRYVSLH--------------------  488  Q6NNA9   Q6NNA9_DROME
469  AHGADRSGKDGVMEMNSIEPAKETTTNV  496  P11169   GTR3_HUMAN
```

Figure 39. Aminoacid sequences alignment of the *Drosophila* Glut1 and GLUT3.

The proteins are shown with their Uniprot code.

Data provided in FlyAtlas, an online database of a comprehensive expression study in *Drosophila*, showed that *Glut1* mRNA is mainly expressed in the brain of the adult fly (http://www.flyatlas.org/ (Chintapalli *et al*. 2007)) (Figure 40). Furthermore, comparing each structural domain and

functional residues in the *Drosophila* Glut1 and the human GLUT1-5 a transmembrane transporter activity can be inferred upon Glut1 (Escher *et al*. 1999). All this evidence and the availability of transgenic strains (for modulating *Glut1* expression) make *Glut1* the most promising candidate for designing a *Drosophila* model to recapitulate observed aberrations in the human *SLC2A3* gene.

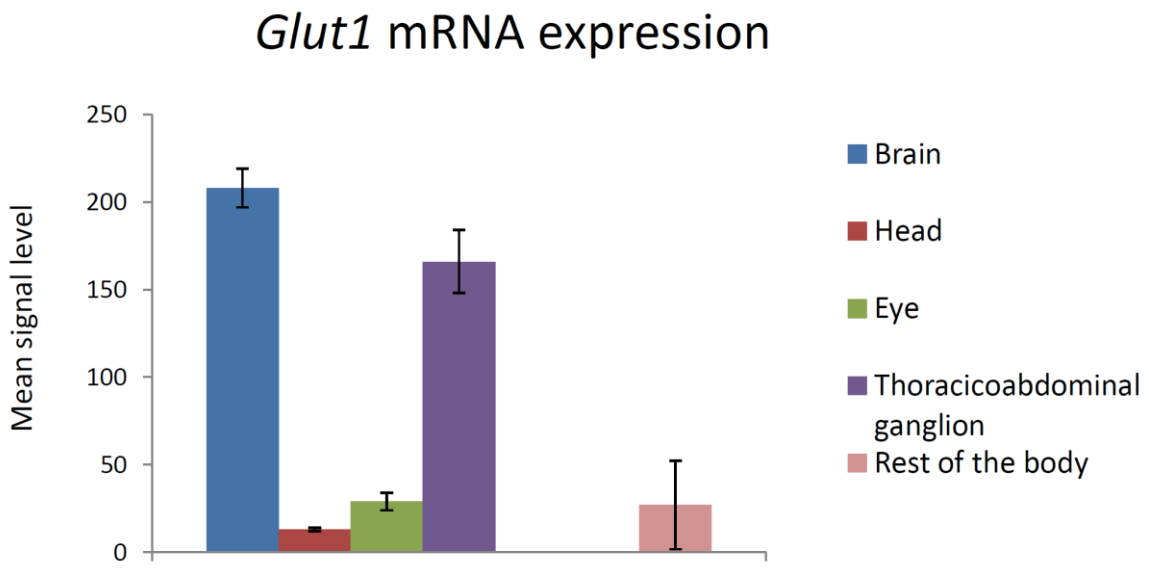


Figure 40. *Drosophila Glut1* mRNA expression profile - array data.

Data from Chintapalli et al. 2007.

Among several *Drosophila* HD models, we used transgenic flies encoding a human HTT exon 1 fragment with a 93 glutamine (Htt93Q) repeat length under control of a *UAS*-enhancer; the expression of the transgene is induced by a *GAL4* driver using the *GAL4/UAS* bitransgenic system (Steffan *et al*. 2001). This model using the pan-neuronal *elavGAL4* driver, allowing selectively the expression of the mutant *HTT* in the *Drosophila* brain, shows phenotype changes attributable to mutant *HTT* insertion such as early toxicity from larva stage, progressive neurodegeneration detectable by the analysis of the pseudopupil as well as reduced half-life.

In order to evaluate the role of *Glut1* in a *Drosophila* HD model (Htt93Q) we tested the effect of reduction, alteration and over-expression of *Glut1* gene expression (RNAi, mutant lines, EP lines) in the HD background. Changes in disease-relevant phenotypes were evaluated using

different metrics such as loss of the rhabdomeres, longevity and eclosion rate. The pseudo-pupil test gives direct information of the neurodegenerative process through counting the loss of visible rhabdomeres of mutant flies. Through analysis of eclosion, which takes in account the number of viable mutant flies versus non mutant flies at the emergence from the pupa, and the longevity, for which the life span of flies of interest are compared, it is possible to deduce if modulation of the gene of interest has beneficial or toxic effect on the fly.

4.4.1 Validation of fly strains used

The expression of *Glut1* mRNA levels in the fruit fly strains carrying a construct for the expression shRNA targeting *Glut1* via RNAi (*w*; *Glut1*^{KK108683}*;+)* and EP construct for the overexpression of *Glut1* (*w*;; *Glut1*^{d05758}*)* was validated by relative qPCR analysis. These strains were crossed with *yw*;; *actinGAL4*/*TM6B* in order to have a F_1 progeny where *actinGAL4* induces the RNAi or overexpression of *Glut1* ubiquitously. The relative expression of *Glut1* in these flies was compared to the driver control (*w*;; *actinGAL4*/*+*). For each strain five biological replicates were used in our investigation. The relative expression of *Glut1* was normalised for *Rpl32* and calculated as ΔCt. ΔCt for each biological replicate was normalised for the mean ΔCt of the controls. ANOVA and multiple comparisons test were performed to test the difference in *Glut1* relative expression between the strains. ANOVA and post-hoc test of comparison with the driver control showed that the relative expression of *Glut1* was decreased by 34% in flies expressing RNAi for *Glut1* (ANOVA, post-hoc test p value = 0.023) and increased by 78% in flies expressing the EP construct for *Glut1* overexpression (ANOVA, post-hoc test p value = 0.00023) (Figure 41). Thus, the genetic manipulations modulating *Glut1* expression have been confirmed.

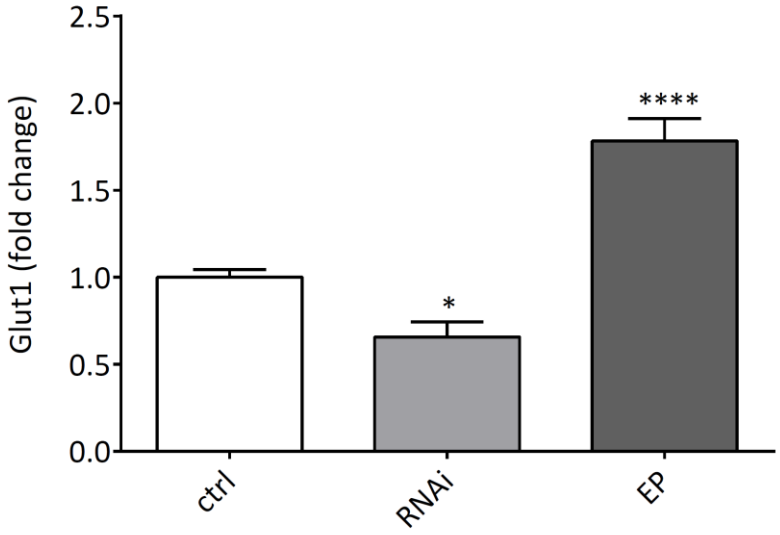


Figure 41. Relative quantification of *Glut1* expression levels in the fly strains used.

*Relative quantification of normalised ΔCt mean (± SEM) per ctrl, RNAi and EP. n≥ 30 flies per each strain. Statistical comparisons by ANOVA and post hoc tests versus ctrl flies. (*p value <0.05; **** p value < 0.0001). "ctrl" refers to yw;; actinGAL4/+; "RNAi" refers to w ; Glut1^{KK108683};actinGAL4; "EP" refers to w;; actinGAL4/Glut1^{d05758}.*

4.4.2 *Glut1* downregulation in HD flies

We initially analysed RNAi-*Glut1*_Htt93Q and the relative control strains (Table 27) for our metrics of investigation: eclosion rate, rhabdomere neurodegeneration and survival rate (see Supplementary material). The same was done for Htt93Q^{3M} and the relative controls strains (Table 27) (see Supplementary material).

4.4.2.1 Eclosion test results

The eclosion rate was calculated for *Glut1-RNAi*_Htt93Q flies from 10 ♂*elavGal4;;* X ♀w; *Glut1^{KK108683}* ;*UAS-Q93httexon1* crosses and the same strategy was used for Htt93Q^{3M} flies (♂*elavGal4;;* X ♀w; *3M*; *UAS-Q93httexon1*). For all of these crosses only the female F_1 progeny expressed the transgenes of interest and the eclosion rate was calculated as total female F_1/ total F_1. The eclosion rate for *Glut1-RNAi*_Htt93Q was stastically different from the eclosion rate of the controls (see Supplementary material) except *elavGAL4;KK; UAS-Q93httexon1* strain, where a KK represents the empty "site specific" transgenic inserting site (ANOVA, post hoc test p value ≥

0.05). As well, the eclosion rate for *Glut1-RNAi*_Htt93Q was not statistically different from the eclosion rate of Htt93Q^{3M} (ANOVA, p value = 0.48) (Figure 42). Therefore, it appeared that the knocking down of *Glut1* in a HD background does not further impair the development of fruit flies.

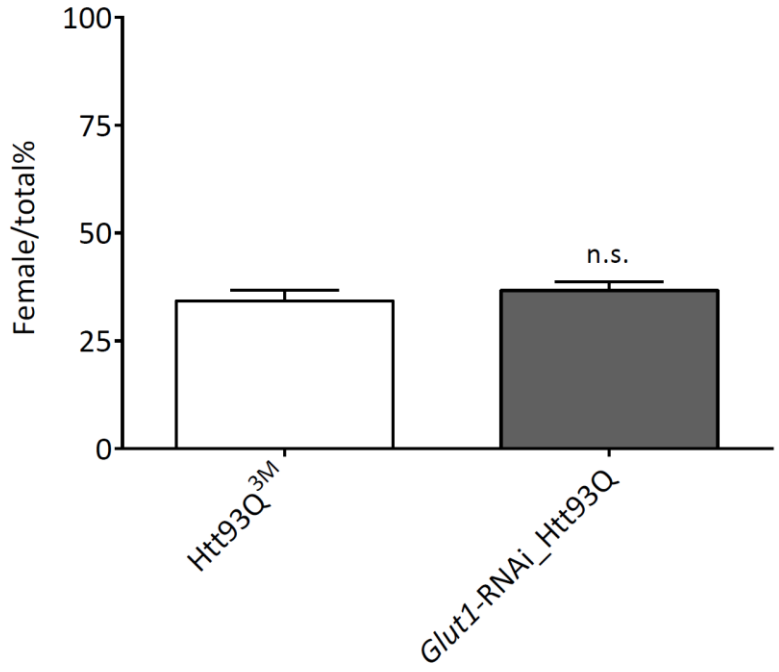


Figure 42. Eclosion rate for Htt93q^{3M} flies and *Glut1*-RNAi_Htt93Q flies.

Knocking down Glut1 expression in flies carrying the mutant HTT does not affect the eclosion rate. n ≥ 1000 progeny for each cross. Statistical comparisons by ANOVA (n.s. not significant).

4.4.2.2 Pseudopupil test results

The loss of rhabdomeres, detected by peudopupil assay, was estimated at day 1 and 7 post-eclosion in *Glut1-RNAi*_Htt93Q and Htt93Q^{3M}. In both *Glut1-RNAi*_Htt93Q flies and Htt93Q^{3M} flies a progressive loss of rhabdomeres was detected; the mean rhabdomeres per each ommatidia analysed was statistically different between *Glut1-RNAi*_Htt93Q flies and Htt93Q^{3M} flies at day 1 (MANOVA, post-hoc test p value = 0.00012) and at day 7 (MANOVA, post-hoc test p value = 0.028) (Figure 43). Thus, it appeared that downregulating *Glut1* by RNAi in a HD background enhanced the neurodegeneration in the adult fly.

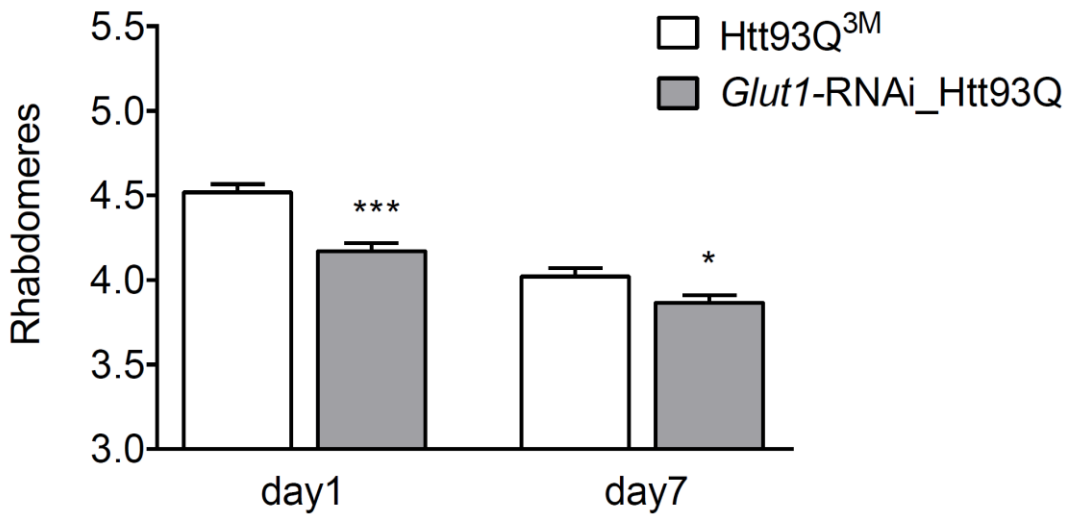


Figure 43. *Glut1*-RNAi_Htt93Q flies exhibit augmented rhabdomere loss compared with Htt93Q^{3M} flies.

*Quantification of mean rhabdomeres (± SEM) per ommatidium in Htt93Q and Glut1-RNAi flies at day 1 and day 7 after eclosion. n= 9 to 15 flies per genotype. Statistical comparisons by MANOVA and post hoc tests versus Htt93Q^{3M} flies (*p value <0.05; *** p value < 0.001).*

4.4.2.3 Lifespan test results

The lifespan of RNAi-*Glut1*_Htt93Q flies was recorded and compared to Htt93Q^{3M} flies. Statistical analysis by Kaplan–Meier survival curve analysis with log rank test was calculated on 100 animals per genotype showing that RNAi-*Glut1*_Htt93Q flies have a decreased lifespan (median lifespan = 17 days) compared to Htt93Q^{3M} flies (median lifespan = 18 days), with a reduction of the median life span of 1 day (p value = 0.034) (Figure 44). Thus, it appeared that reducing levels of *Glut1* by RNAi modestly exacerbated HD phenotype in the adult fly.

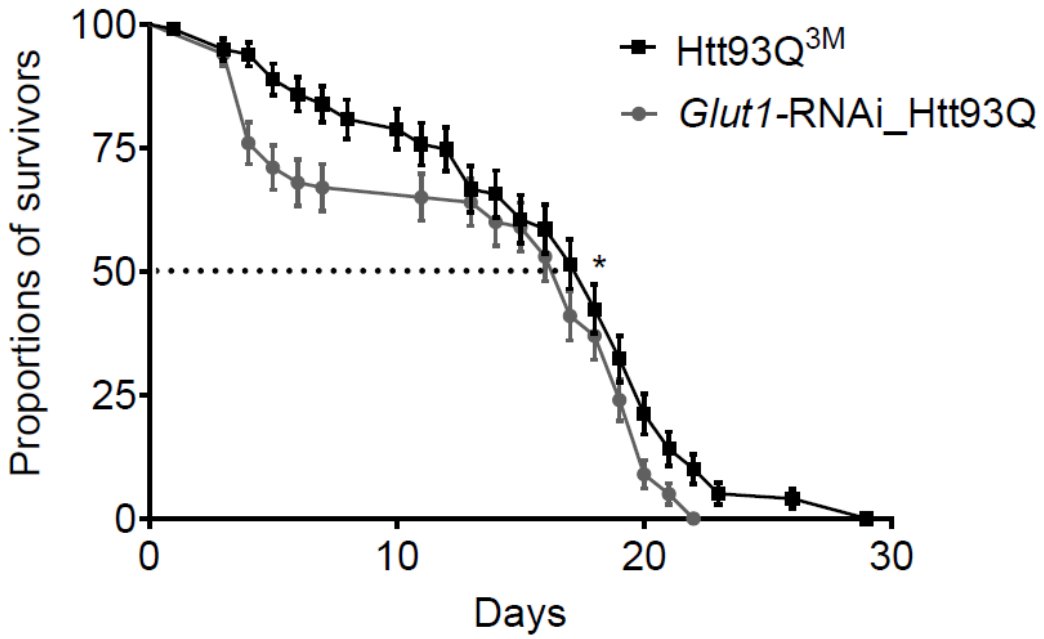


Figure 44. *Glut1*-RNAi_Htt93Q flies exhibit reduced life span compared with Htt93Q^{3M} flies.

*Glut1 downregulation decreased median survival of adult HD flies. The number of flies surviving from each cohort was determined every day. Statistical analysis by Kaplan–Meier survival curve analysis with log rank test (*p value < 0.05). n=100 animals per genotype.*

4.4.3 *Glut1* overexpression and Glut1 mutant in HD flies

We initially analysed EP-*Glut1*_Htt93Q, mut-*Glut1*_Htt93Q and their relative control strains (Table 27) for our metrics of investigation: eclosion rate, rhabdomere number and lifespan(see Supplementary material). The same was done for Htt93Q and the relative controls strains (Table 27) (see Supplementary material). Afterwards, EP-*Glut1*_Htt93Q and mut-*Glut1*_Htt93Q were compared to Htt93Q for our metrics of investigation.

4.4.3.1 Eclosion test results

The eclosion rate for EP-*Glut1*_Htt93Q was calculated from 10 ♂*elavGal4;;* X ♀w; *Glut1*^{d05758} ;*UAS-Q93httexon1* crosses, for mut-*Glut1*_Htt93Q was calculated from 10 ♂*elavGal4;;* X ♀w; *Glut1*^{17J} ;*UAS-Q93httexon1* crosses and for Htt93Q flies from 10 ♂*elavGal4;;* X ♀w;; *UAS-Q93httexon1* crosses. For all of these crosses the ratio of the female F_1 progeny, expressing the

transgenes of interest, versus the total F_1 progeny was counted. ANOVA of the eclosion rate of EP-*Glut1*_Htt93Q, mut-*Glut1*_Htt93Q and Htt93Q showed that there was a significant correlation between eclosion rate and genotype (ANOVA, p value = 0.0048). Mut-*Glut1*_Htt93Q flies was significantly associated with a reduced eclosion rate compared to Htt93Q flies (ANOVA, post hoc test, p value = 0.015) and EP-*Glut1*_Htt93Q flies (ANOVA, post hoc test, p value = 0.0023) (Figure 45). Thus, it was clear that mutation of *Glut1* can worse early disease events in a fly model of HD and overexpression of *Glut1* can potentially rescue from this process, though the increase in eclosion rate between Htt93Q and EP-*Glut1*_Htt93Q flies was not significant.

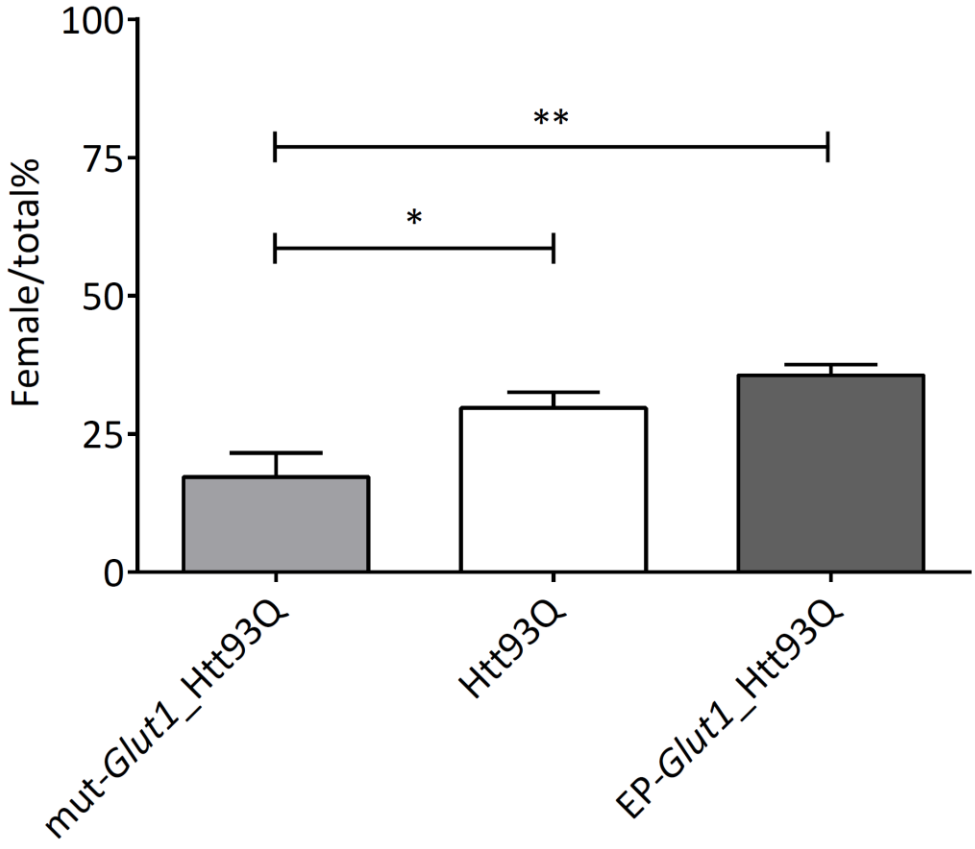


Figure 45. Alterations of *Glut1* affect eclosion rate in HD background.

*Mut-Glut1_Htt93Q flies showed a decreased adult emergence from the pupal case compared to Htt93Q and EP-Glut1_Htt93Q flies. n ≥ 1000 progeny for each cross. Statistical comparisons by ANOVA and post-hoc test (*p value < 0.05; ** p value < 0.01).*

4.4.3.2 Pseudopupil test results

The pseudopupil test was performed on mut-*Glut1*_Htt93Q flies and EP-*Glut1*_Htt93Q flies compared to Htt93Q flies at day 1 and day7 post eclosion. EP-*Glut1*_Htt93Q flies showed a significant reduction of neuronal loss at day 1 and day 7 compared to Htt93Q flies (MANOVA, post-hoc test, p value = 0.0011 at day 1, p value = 0.0013 at day 7) as well to mut-*Glut1*_Htt93Q flies (MANOVA, post-hoc test, p value = 0.00016 at day 1, p value = 0.0037 at day 7) (Figure 46). From these data it was possible to say that overexpressing *Glut1* partially rescued neurodegeneration in adult HD flies and alteration of *Glut1* could potentially worsen the neurodegenerative process, although the decrease in rhabdomere number between Htt93Q and mut-*Glut1*_Htt93Q flies was not significant.

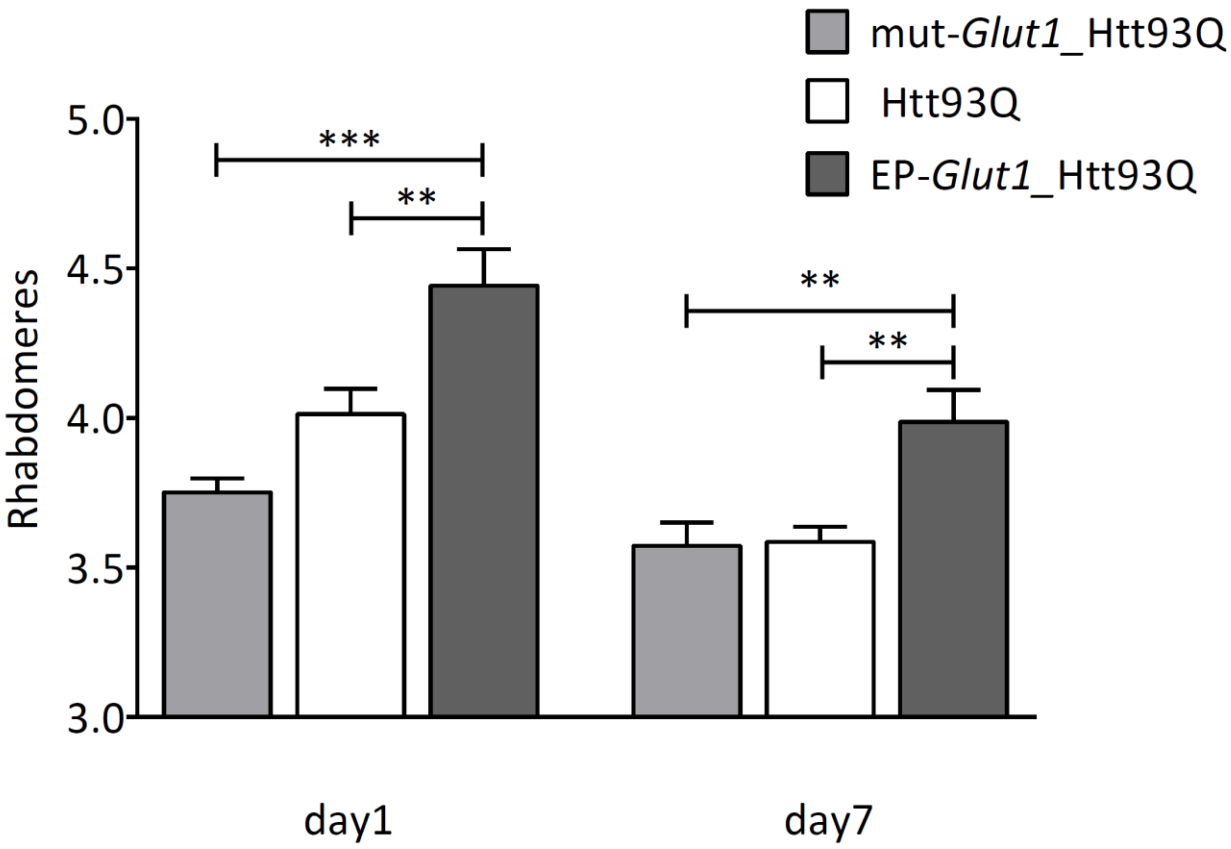


Figure 46. Overexpression of *Glut1* reduces neuronal loss in HD flies.

Quantification of mean rhabdomeres (± SEM) per ommatidium in Htt93Q, mut-*Glut1_Htt93Q* and EP-*Glut1_Htt93Q flies at day 1 and day 7 post eclosion. n= 9 to 15 flies per genotype. Statistical comparisons by MANOVA and post hoc tests (**p value < 0.01; *** p value < 0.001).*

4.4.3.3 Lifespan test results

The lifespan of mut-*Glut1*_Htt93Q flies and EP-*Glut1*_Htt93Q flies was recorded and compared to Htt93Q flies. Kaplan–Meier survival curve statistical analysis with log rank test was calculated on 100 animals per genotype and it showed that the survival rate of Htt93Q flies was not significantly different compared to mut-*Glut1*_Htt93Q flies (p value = 0.11) (Figure 47) and EP-*Glut1*_Htt93Q flies (p value = 0.52) (Figure 48). Thus, in this experimental context, alteration or overexpression of *Glut1* did not affect the lifespan in HD model flies.

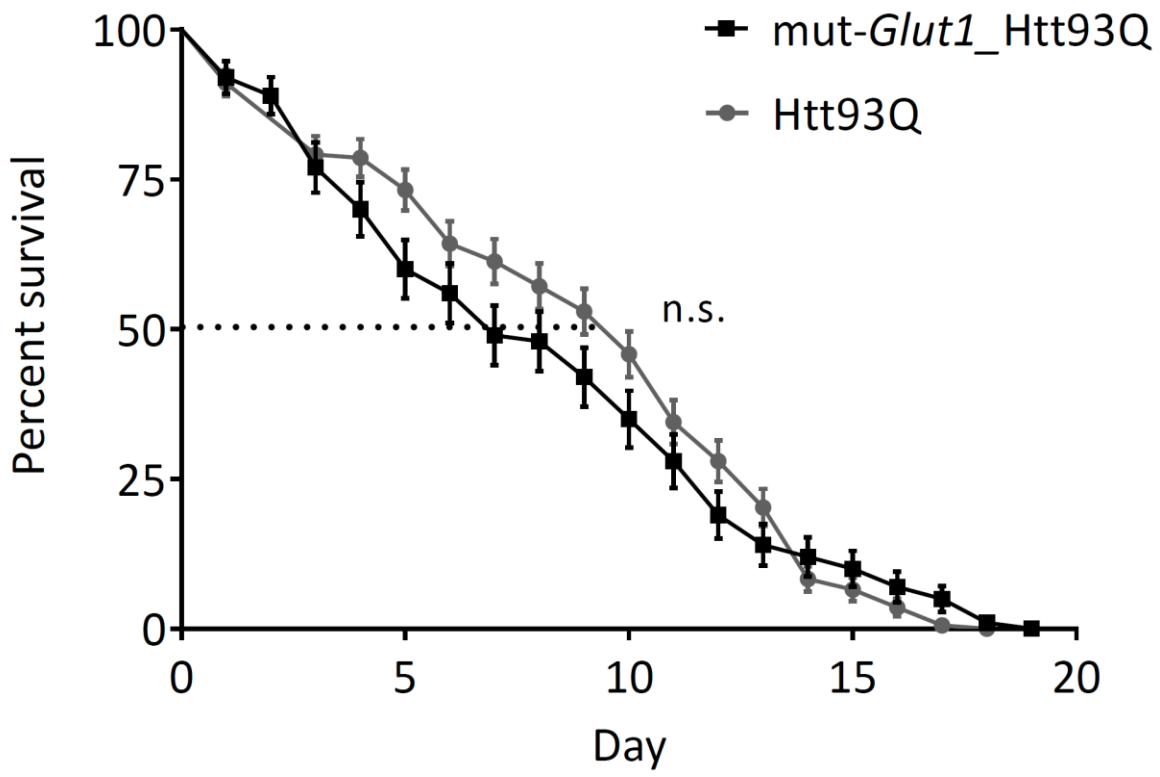


Figure 47. mut-Glut1_Htt93Q flies exhibit no different life span compared with Htt93Q flies.

Glut1 alteration does not affect survival of adult HD flies. The number of flies surviving from each cohort was determined every day. n=100 animals per genotype. Statistical analysis by Kaplan–Meier survival curve analysis with log rank test (n.s. not significant).

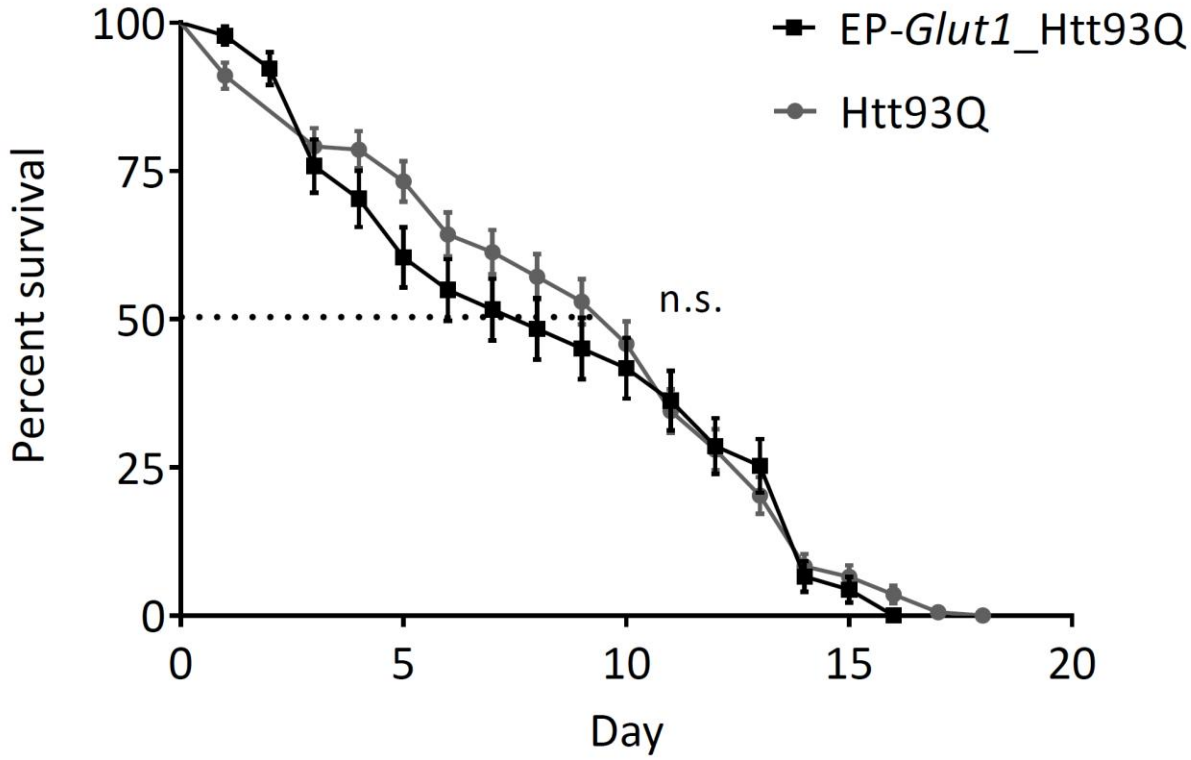


Figure 48. EP-*Glut1*_Htt93Q flies exhibit no different life span compared with Htt93Q flies.

Glut1 overexpression does not affect survival of adult HD flies. The number of flies surviving from each cohort was determined every day. n=100 animals per genotype. Statistical analysis by Kaplan–Meier survival curve analysis with log rank test (n.s. not significant).

CHAPTER FIVE: DISCUSSION

Encompassing nearly 4 % of the human genome and leading to structural rearrangements that involve from ~ 50bp up to several Mb, CNV is a large contributor to genomic variation. CNVs have been investigated with regards to their functional impact on the full range of biology, from gene-expression studies to GWAS interrogating all the classes of human genetic diseases: Mendelian, complex, sporadic, infectious and neurodegenerative.

HD is a monogenic disorder caused by an autosomal dominant mutation in the *HTT* gene. Notwithstanding its relentless course, the wide temporal range of AO (ranging from 1 year to 80 years) of the motor symptoms and the several phenotypes associated with HD are not fully accounted for by the expanded CAG repeat in the *HTT* gene. As previously shown in a study on Venezuelan families, the additive genetic heritability of the residual age of onset is 38%, suggesting the presence of other genetic factors involved which modulate onset and progression of the disease (Wexler et al. 2004). Several studies have analysed common genetic variants in manifest HD carrier cohorts to test whether they are associated with earlier or later AO among similar CAG length carriers. Notwithstanding, few genetic modifiers of the AO so far identified have been confirmed by independent studies. Beyond the genetic variability of the candidate modifiers studied, it is fundamental to consider the heterogeneity and complexity of HD pathology. The association of genetic variants to phenotypic variance in HD, within or without taking into account the effect of the HD mutation, needs to be assesed on a large HD sample cohort that can widely represent multiple aspects of HD pathology in terms of CAG mutation length, age of onset, sex, major symptoms at onset and ethnicity. As such information on a large panel of HD patients are provided by international consortia such as the EHDN, it is important to standardize the practise of evaluation for candidate genetic modifier of the AO of HD with regards of the characteristics of the sample cohort adopted and appropriate statistical methods for the analysis. Additionally, accessible online databases should be created to collect all the data from genetic analysis performed on each HD patient, as well as clinical assessment data of HD biomarkers evaluation, psychiatric, cognitive and motor test scores were accessible.

To our knowledge this is the first study that tests the potential role of CNVs (e.g. *DEFB4* and *SLC2A3*) as genetic modifiers of the age of onset in HD. This study relies on the application of a validated PRT method for CNV genotyping (Aldhous *et al.* 2010; Veal *et al.* 2013) and investigates the impact of CNV on the AO in HD on large sample cohorts which accurately reflect the HD

population in terms of CAG mutation length, AO, sex,major symptoms at onset and sample provenence.

5.1 β-defensin CNV in HD

From our study it is clear that β-defensin genomic copy number is not associated with the modulation of HD pathogenesis and does not affect the age of onset of any main symptoms at onset. Nevertheless, we cannot formally exclude the potential implication of single nucleotide polymorphisms within this region, given that of course not only copy number change but also sequence variants can affect the expression of hBD2 (Groth *et al*. 2010).

In order to explain the output of our investigation, firstly it is important to take into account the characteristic of the gene-dosage effect of β-defensin CNV on the expression of h-BD2. So far a number of studies showed the correlation between genomic copy number and h-BD2 mRNA or protein level in several tissues. A strong correlations (R^2 = 0.96; p = 0.02) between β-defensin copy number and h-BD2 mRNA level was found in nasal epithelial cells (Janssens *et al*. 2010); higher significances but moderate correlations between genomic copy number and protein level were found in serum of healthy (Pearson's R = 0.46 p< 7×10^{-5} in Jansen *et al*. 2009; Pearson's r correlation = 0.370; p < 0.001 in Jaradat *et al*. 2013) and disease individuals (Pearson's r correlation = 0.458; p < 0.001 in Jaradat *et al*. 2013). These studies suggest that the effect of β-defensin CNV on h-BD2 gene expression have a stonger effect in cell type where h-BD2 is normally expressed rather than secreted, and might vary in normal or a disease background. Beyond that, it is important to say that I took in account as evidence of the dosage effect mechanism only Jansen *et al*. 2009 and Hollox *et al*. 2003, based on reliable and more accurate methods for the CNV genotyping have been used (qPCR: Janssens *et al*. 2010 and Jaradat *et al*. 2013; PRT and REDVR, and MAPH: Jansen *et al*. 2009; MAPH, FISH, PFGE: Hollox *et al*. 2003).

With regards to our hypothesis, we can say that the activated microglia and reactive astrocytes, shown to be correlated with disease severity in HD, might not influence (or be influenced by) the β-defensin CNV. The inducible expression of hBD2 in astrocytes driven by TNF-α

and IL-β (Hao *et al*. 2001), which are both expressed in the striatum of HD brains (Björkqvist *et al*. 2008), or by the aberrant activation of NF-κB (Hsiao *et al*. 2013), may not be copy number dependent or the gene-dosage effect could be responsible for a weak correlation with protein level in astrocyte, negligible in the HD brain. It is also possible that transcriptional and trafficking alterations in astrocytes due to mutant huntingtin (Kuhn *et al*. 2007; Luthi-Carter and Cha 2003) might interfere with hBD2 expression or function, masking any possible modulation in the expression driven by its copy number. Furthermore, we can say that inflammatory stimuli at the BBB might not activate hBD2 expression contributing significantly to the inflammatory burden in HD and further studies are actually necessary to clarify the role of hBD2 on the BBB. This study does not exclude any possible implication of β-defensins in HD which could contribute to a pro-inflammatory state within the brain or exacerbating the inflammatory response (Williams WM *et al*. 2012) in neurodegenerative disorders. Further studies are necessary to clarify the effects of β-defensin copy number on hBD2 expression in CNS tissues in both normal physiological and neurodegenerative milieus.

5.2 *SLC2A3* CNV in HD

From our genetic screening and analysis in patient cell lines, it is possible to say that *SLC2A3* CNV is a genetic modifier of the age of onset in HD. We determined *SLC2A3* copy number for 988 HD individuals, representing an extensive sample cohort, appropriate for the detection of genetic modifiers of the AO in HD.

In order to test any association between *SLC2A3* CNV distribution and variance of the AO in HD, we used the natural-logarithmic transformed AO into a generalized linear model, where the CAG mutation length and the *SLC2A3* CNV were independent predictors. Using the log-transformed AO improves the linearity of the regression between the dependent variable and the CAG mutation length which is its main predictors. Several studies used a simple linear regression model in which it is erroneously assumed a constant, symmetrical variance of onset ages around the estimated means. Instead, the using of a GLM betters to address linearity in response on the specified scale and accommodates skewness through variance weighting.

From the analysis of the *SLC2A3* CNV distribution in our two HD cohorts it was evident that *SLC2A3* CNV is a rare genomic variation event, involving < 5% of the population. The frequencies of *SLC2A3* CNV distribution in the two cohorts were not significantly different although the analysis of their correlation with variance of AO in the separate cohorts gave contradictory results. More likely slight differences between the two cohorts (for instance in HD cohort 2 there are 5% more of samples with AO > 60 years old than in HD cohort 1, which could have affected the correlation between CAG and AO and therefore the size effect of *SLC2A3* CNV on the AO prediction) and stochastic effect on *SLC2A3* copy number distribution in the two cohorts significantly increased and reduced the power of the effect of *SLC2A3* CNVs respectively in HD cohort 1 and HD cohort 2. However, considering that two cohorts are samples of the HD population, it was possible to estimate the effect of *SLC2A3* CNV combining the two cohorts together in order to have a larger samples set. From the analysis of the effect of SLC2A3 CNV on the AO in 988 HD samples it arises that increase in copy number is significantly associated to a delay of up to ~ 6 years of AO in HD, a sizeable effect comparable to another genetic modifier of AO such as *HAP1*, where the protective allele is associated to a delay of the AO of 8 years.

We investigated the functional effect of *SLC2A3* CNV analysing the GLUT3 expression level in lymphoblastoid cell lines of previously genotyped patients from our HD cohorts. A significant correlation between *SLC2A3* duplication and an augmented GLUT3 expression compared to samples with 1 and 2 *SLC2A3* copies was found. The functional effect on protein expression levels allowed us to indentify *SLC2A3* CNV as a genetic modifier of the AO of HD. While *SLC2A3* duplication, associated with an increased gene expression, was found to be protective, *SLC2A3* deletion was not associated to earlier onset of HD. Likely no detrimental effect was visible due to the small number of 1 *SLC2A3* copy carriers in our cohorts (n=5). Regarding the analysis of GLUT3 relative expression level in patient cell lines, no significant difference was found between *SLC2A3* deletion carriers and controls probably due to a stochastic effect involving protein expression related to low gene copies (Elowitz *et al.* 2002).

In order to clarify the role of *SLC2A3* CNV in HD, we used a *Drosophila* model of HD to investigate the effects of overexpression and underexpression of a main glucose transporter in the fly CNS. In HD flies the overexpression of the *Drosophila* Glut1, homologue of human GLUT3, rescued neurodegeneration whilst alteration of this transporter was associated with a worsening

of the disease progression. We showed underexpression of Glut1 worsened the neurodegeneration in early stage causing more severe rhabdomere loss at day 1 than at day 7, as well as increased mortality at a few days post eclosion, and an overall reduced eclosion rate. These findings indicate that dysfunctions of Glut1 in a *Drosophila* model of HD are not well tolerated at early stage, likely because of the role of Glut1 during larva development. In agreement with previous studies investigating the role of Glut1 in fly models of HD and AD (Besson et al. 2010; Shulman et al. 2011), the overexpression of Glut1 can rescue the neurodegeneration significantly in the adult fly and with a non significant but indicative effect even in the early stage of the fly life, confirming that higher level of glucose transporter can delay HD progression.

This is the first study to identify a CNV event as a genetic modifier of the age of onset in HD. Specifically we showed that *SLC2A3* CNV, affecting the gene expression level, was associated to a delay of AO in HD. This study highlighted a new aspect of the role of GLUT3 and the neuronal glucose metabolism in HD pathology. As GLUT3 (and GLUT1) reduction was suggested to be involved in HD pathology and the brain hypometabolism, present without a significant atrophy, is a potential biomarker of HD progression in presymptomatic patients, our study supports the hypothesis that increased expression of GLUT3 could counter an early energetic deficit in HD neurons delaying the disease progression (Gamberino and Brennan 1994; Ciarmiello *et al.* 2012). In HD neurons, expression of the additional *SLC2A3*, constitutively expressed or activated by inflammatory or energy-deficiency stimuli, could contrast to the deficit of the transporter, which expression or trafficking could be directly or indirectly affected by the mutant HTT. Further evidences supporting our hypothesis were provided using a *Drosophila* model of HD where we selectively studied the effect of the overexpression and underexpression of the main glucose transporter in the brain in an HD background, causing respectively amelioration and worsening of the disease progression.

To unequivocally confirm the correlation between *SLC2A3* CNV and a variance of the AO in HD, an independent replication study is necessary on an even larger cohort to confirm the size of effect, and to test the effect of deletion. To understand if *SLC2A3* is involved at an early stage in HD progression, it will be fundamental to investigate if regulation of gene expression is affected by the mutant protein. Studies on murine neuroblasts and trophoblasts highlighted the role of Sp1/Sp3 and CREB transcription factors on GLUT3 expression, whose functions are both affected in

HD, suggesting a potential effect on GLUT3 expression dependent upon mutant HTT (Rajakumar *et al.* 1998; Rajakumar *et al.* 2004; Cha 2007). Moreover, dysfunction in vesicle trafficking in HD neurites might affect the recycling of GLUT3, as it colocalizes with SNARE complex proteins that are found to be lowered in HD brains (Heather West Greenlee *et al.* 2003; Morton *et al.* 2001; Smith *et al.* 2007).

To confirm the findings from our in vivo study using the Drosophila HD model, a HD model in mouse bearing an extra copy of Slc2a3,the murine homologue of human *SLC2A3*, should be developed to test if the overexpression of this transporter will be protective against HD and how this will affect the cerebral metabolism and atrophy. From this mouse model and from neuronal cell lines cultured from it, it will be fundamental to analyse any possible effect of GLUT3 overexpression in HD. These studies could represent a further step to explain the role of GLUT3 and specifically of *SLC2A3* CNV in HD and improve our knowledge of HD pathology.

In parallel to these studies screening of xenobiotic compounds able to stimulate Glut1 expression in a *Drosophila* model of HD could be performed in order to define a new therapeutic strategy against the disease. Strikingly, studies in rodent models of aging and Alzheimer's disease have already identified a potential compound (lipoic acid) that could be administered to murine models of HD in order to stimulate GLUT3 expression and rescue the glucose uptake in the brain (Jiang *et al.* 2013; Sancheti *et al.* 2013), and test for disease-related phenotypic improvements.

This study remarks the important role of the glucose metabolism in HD, which alteration is an early biomarker the disease progression, and highlighted a potential therapeutical target, *SLC2A3*, which increased expression could restore the glucose homeostasis in the brain and delay HD progression.

ANNEX

Supplementary material

Eclosion test in control strains.

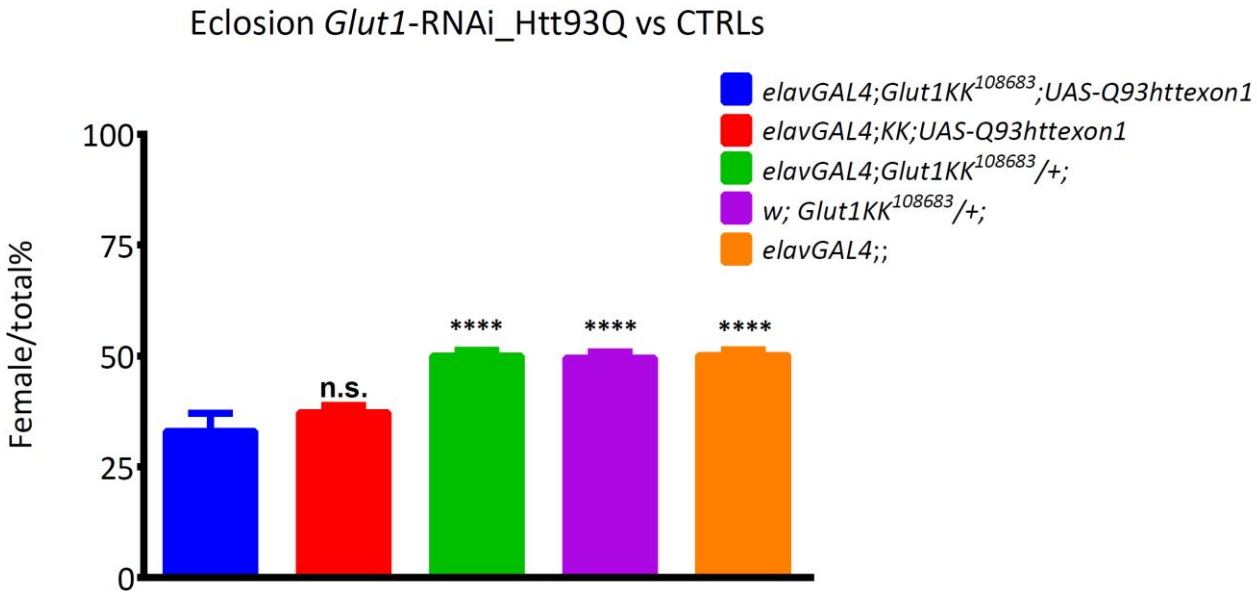


Figure 49. Eclosion rate in Glut1-RNAi and relative control strains.

*Statistical comparisons of mean eclosion rate ((Female F_1/total F_1) %) (±SEM) by ANOVA and post-hoc test versus Glut1-RNAi flies (n.s. not significant; **** p value < 0.0001). n ≥ 1000 progeny for each cross.*

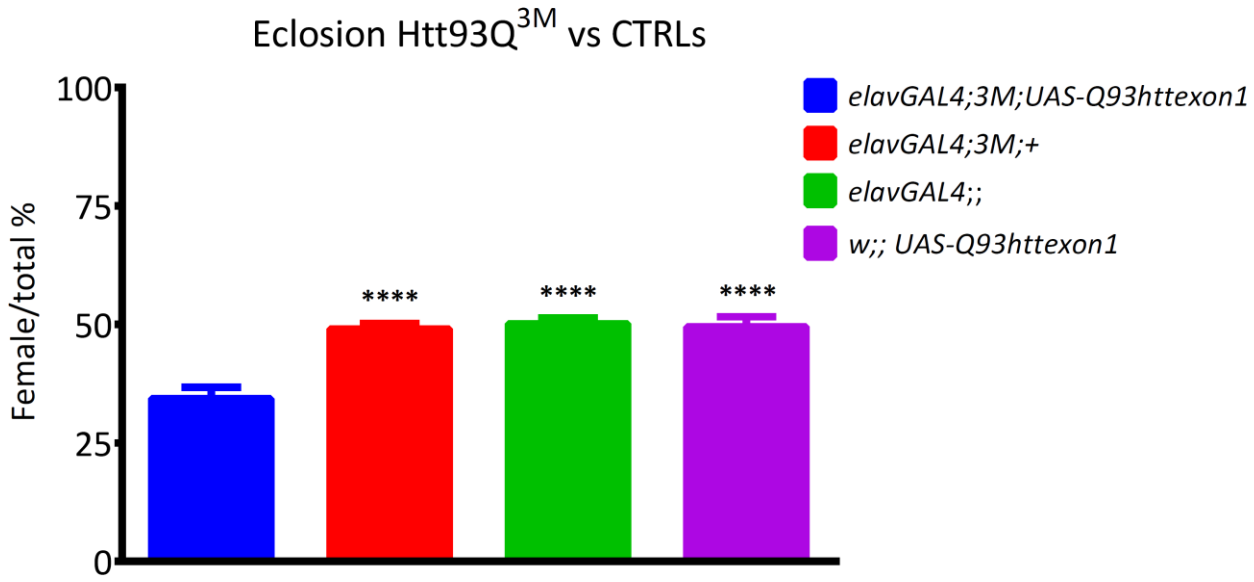


Figure 50. Eclosion rate in Htt93Q^{3M} and relative control strains.

*Statistical comparisons of mean eclosion rate ((Female F_1/total F_1) %) (±SEM) by ANOVA and post-hoc test versus Htt93^{3M} flies (**** p value < 0.0001). n ≥ 1000 progeny for each cross.*

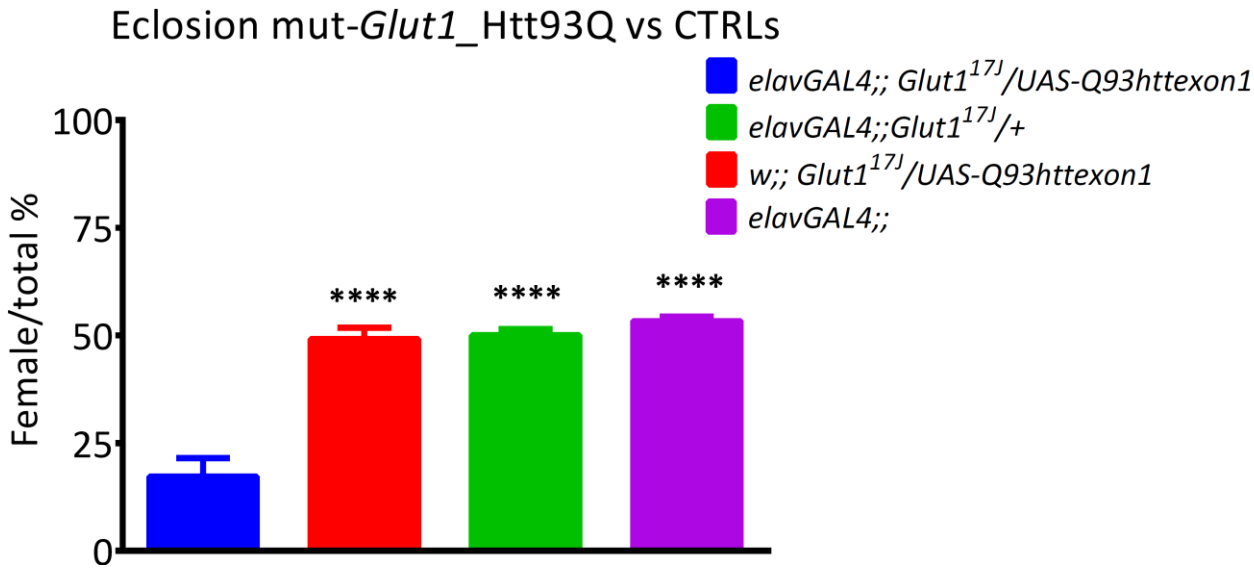


Figure 51. Eclosion rate in mut-Glut1 and relative control strains.

*Statistical comparisons of mean eclosion rate ((Female F_1/total F_1) %) (±SEM) by ANOVA and post-hoc test versus mut-Glut1 flies (**** p value < 0.0001). n ≥ 1000 progeny for each cross.*

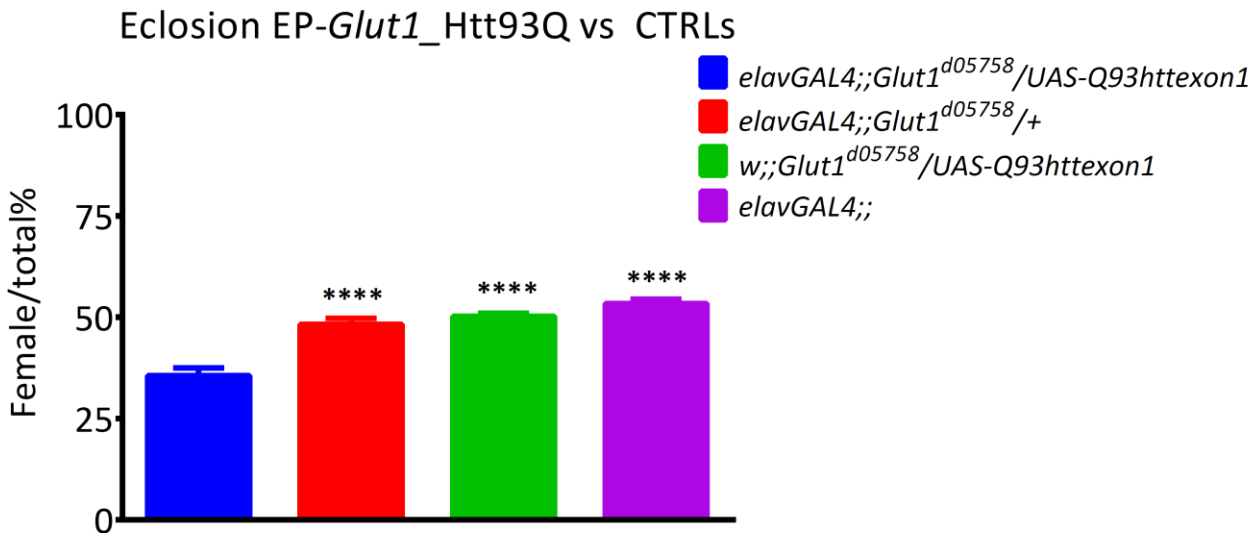


Figure 52. Eclosion rate in EP-*Glut1* and relative control strains.

*Statistical comparisons of mean eclosion rate ((Female F_1/total F_1) %) (±SEM) by ANOVA and post-hoc test versus EP-Glut1 flies (**** p value < 0.0001). n ≥ 1000 progeny for each cross.*

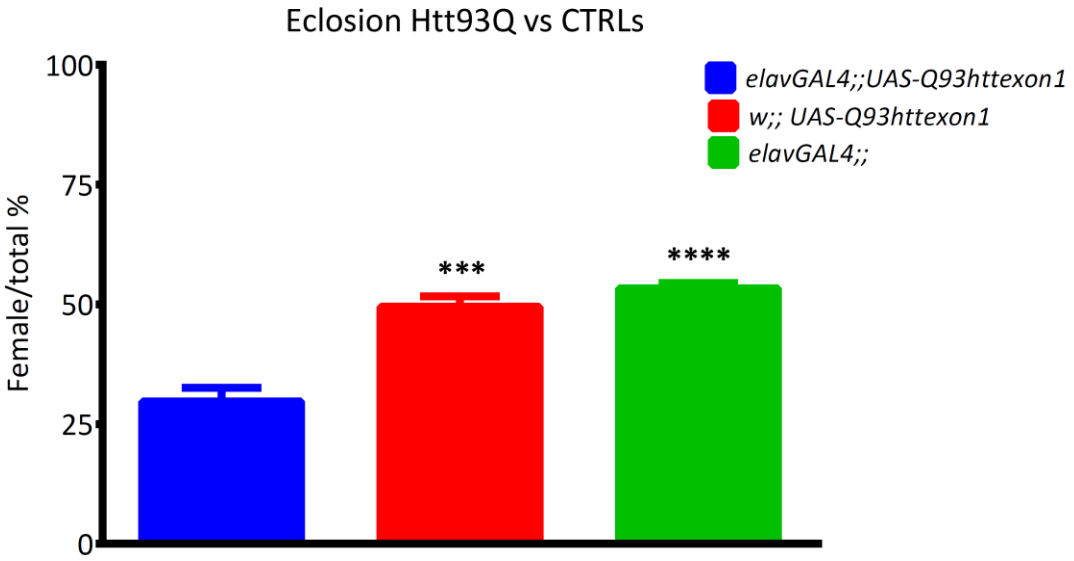


Figure 53. Eclosion rate in Htt93Q and relative control strains.

*Statistical comparisons of mean eclosion rate ((Female F_1/total F_1) %) (±SEM) by ANOVA and post-hoc test versus Htt93Q flies (***p value < 0.001 ;**** p value < 0.0001). n ≥ 1000 progeny for each cross.*

Lifespan test in control strains.

The number of flies surviving from each cohort was determined every day for flies in HD background and every 3 to 4 days for controls strains. Statistical analysis was done by Kaplan–Meier survival curve analysis with log rank test on approximately 100 animals per genotype.

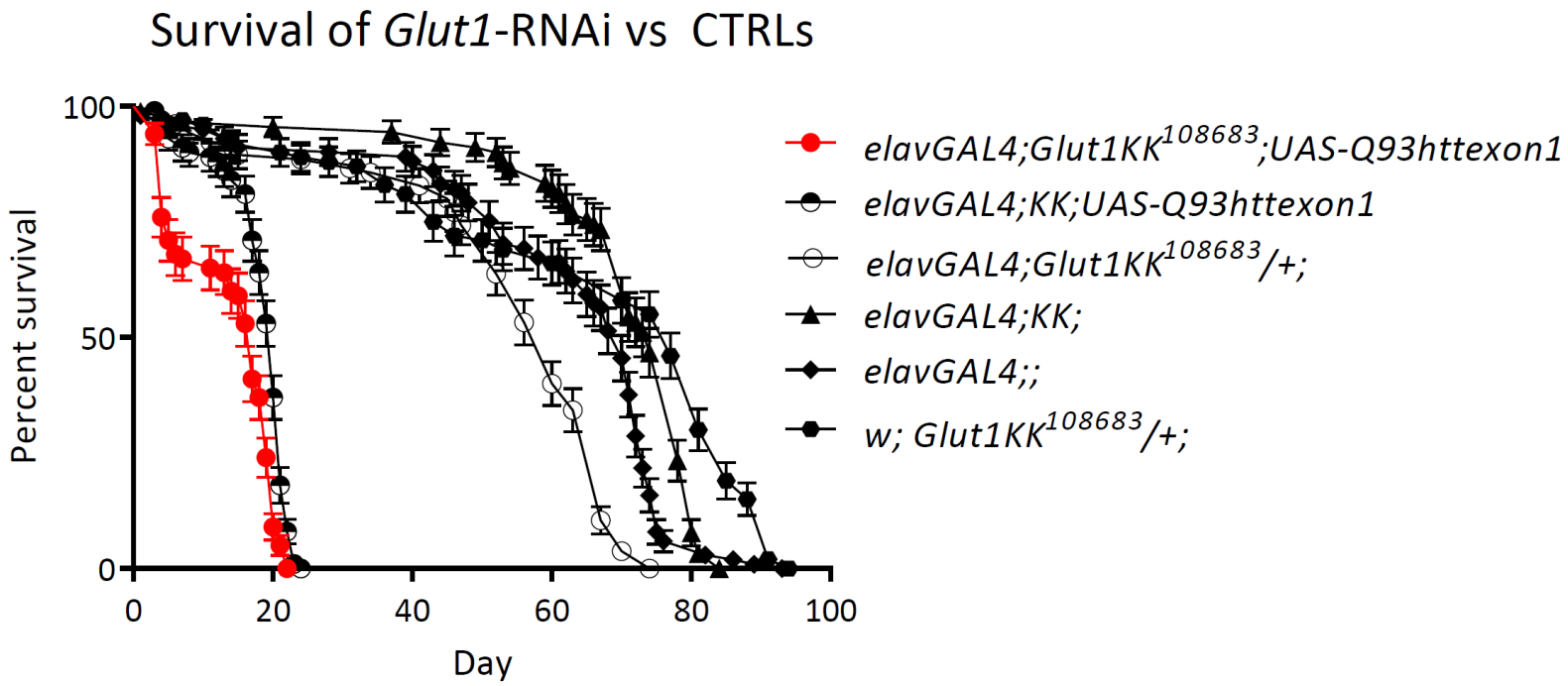


Figure 54. Survival curve of Glut1-RNAi versus control strains.

Table 20. Statistical analyses of survival curve of Glut1-RNAi versus control strains.

Genotype	*elavGAL4;Glut1KK^{108683};UAS-Q93httexon1*	*elavGAL4;KK;UAS-Q93httexon1*	*elavGAL4;Glut1KK^{108683}/+;*	*elavGAL4;KK;*	*elavGAL4;;*	*w; Glut1KK^{108683}/+;*
Median survival (day)	17	20	60	74	70	77
Curve comparison with *Glut1*-RNAi: p value	/	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

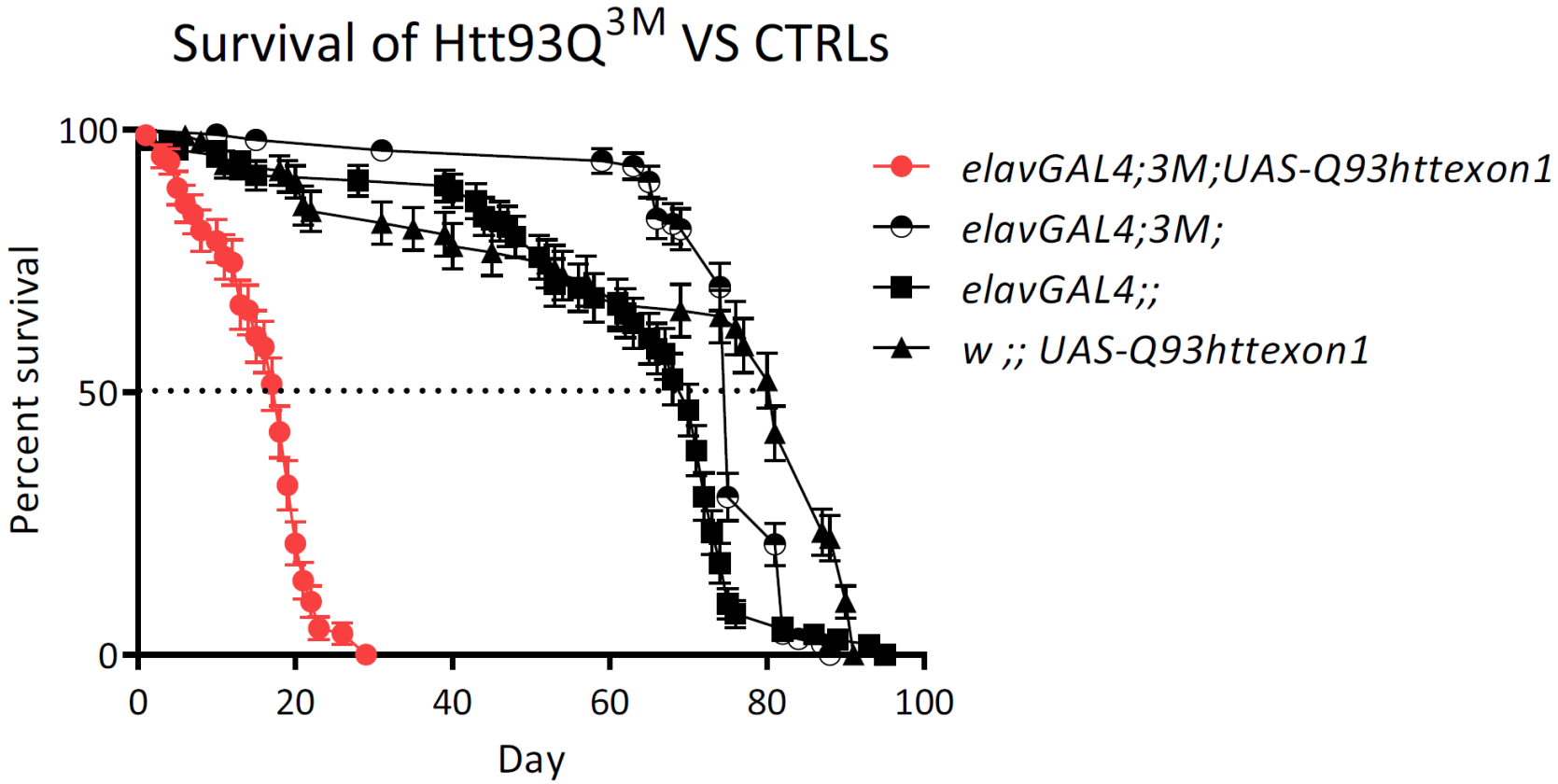


Figure 55. Survival curve of Htt93Q^{3M} versus control strains.

Table 21. Statistical analyses of survival curve of Htt93Q^{3M} versus control strains.

Genotype	*elavGAL4;3M;UAS-Q93httexon1*	*elavGAL4;3M;*	*elavGAL4;;*	*w ;; UAS-Q93httexon1*
Median survival (day)	18	75	70	81
Curve comparison with Htt93Q^{3M} : p value	/	<0.0001	<0.0001	<0.0001

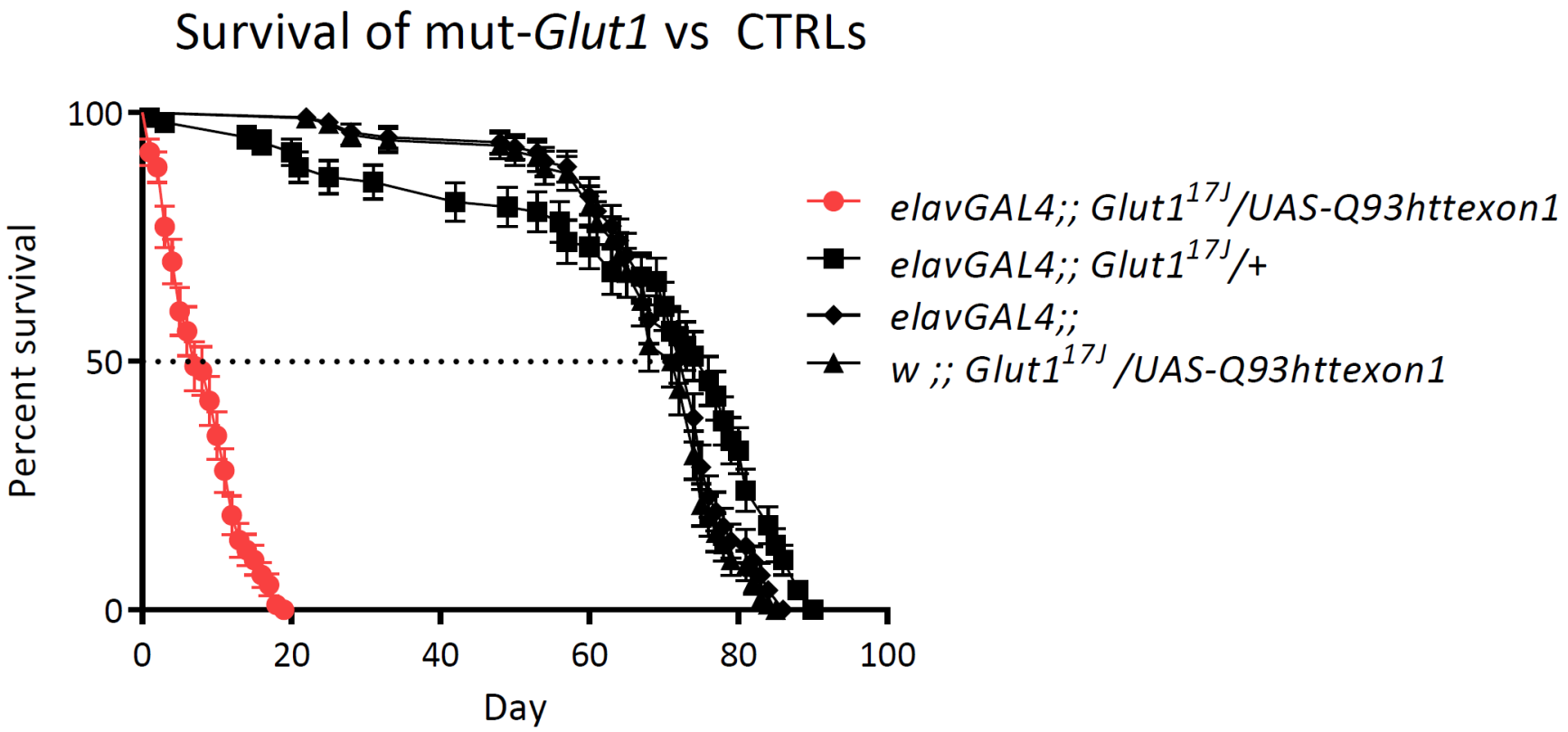


Figure 56. Survival curve of mut-Glut1 versus control strains.

Table 22. Statistical analyses of survival curve of mut-*Glut1* versus control strains.

Genotype	*elavGAL4;; Glut1^{17J}/UAS-Q93httexon1*	*elavGAL4;; Glut1^{17J}/+*	*elavGAL4;;*	*w ;; Glut1^{17J} /UAS-Q93httexon1*
Median survival (day)	7	76	70	71.5
Curve comparison with mut-*Glut1*: p value	/	<0.0001	<0.0001	<0.0001

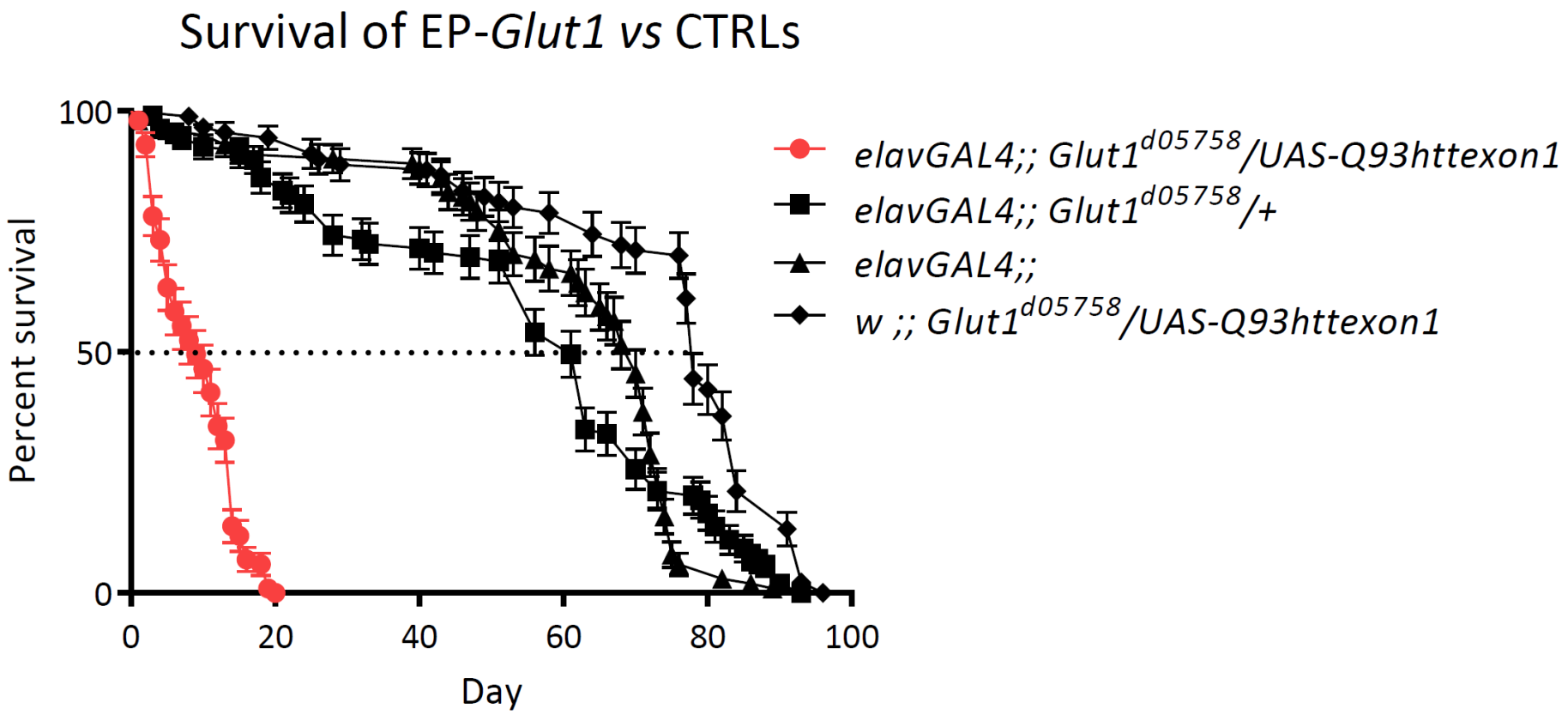


Figure 57. Survival curve of EP-Glut1 versus control strains.

Table 23. Statistical analyses of survival curve of EP-*Glut1* versus control strains.

Genotype	*elavGAL4;; Glut1^{d05758}/UAS-Q93httexon1*	*elavGAL4;; Glut1^{d05758}/+*	*elavGAL4;;*	*w ;; Glut1^{d05758}/UAS-Q93httexon1*
Median survival (day)	9	61	70	78
Curve comparison with EP-*Glut1* : p value	/	<0.0001	<0.0001	<0.0001

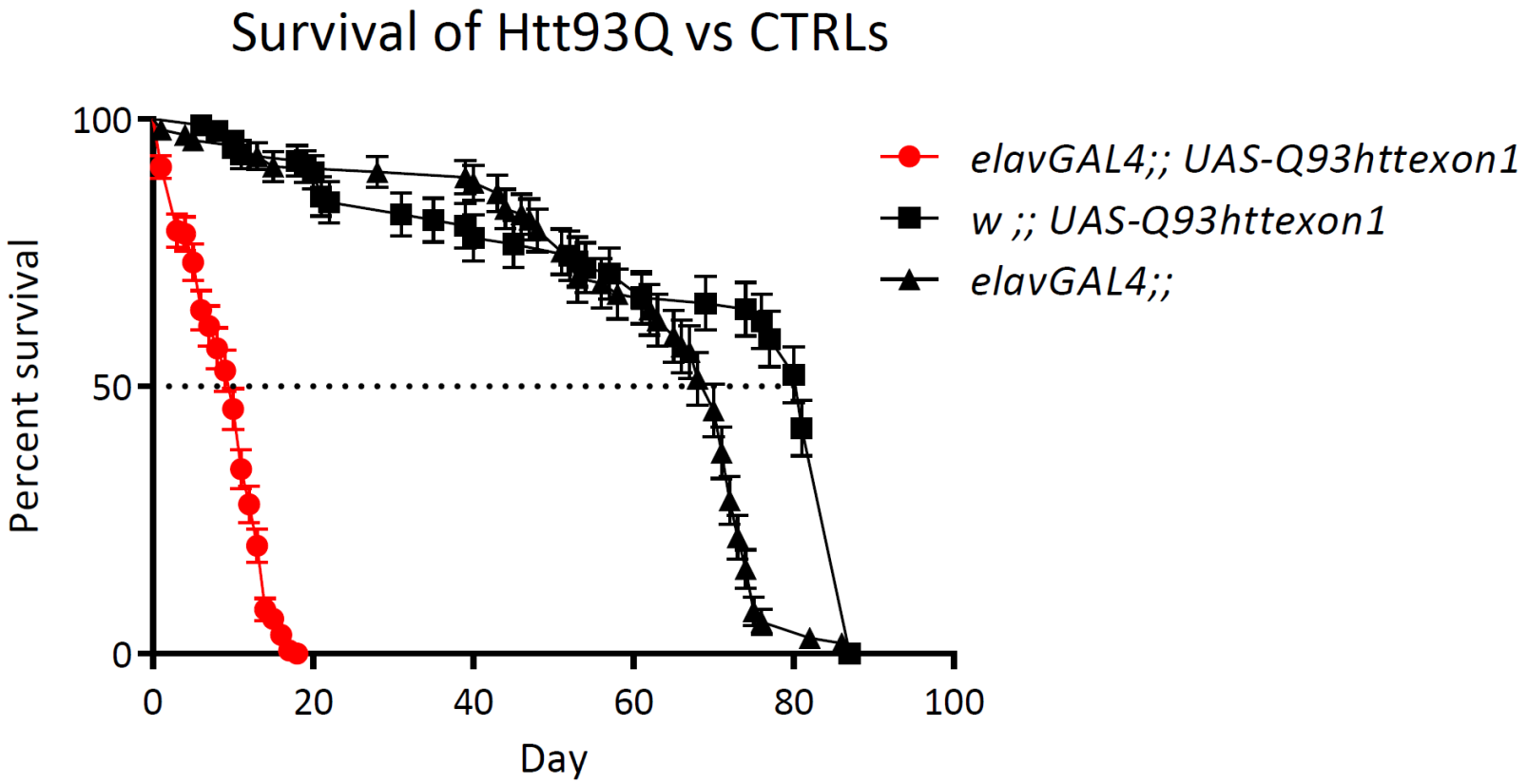


Figure 58. Survival curve of Htt93Q versus control strains.

Table 24. Statistical analyses of survival curve of Htt93Q versus control strains.

Genotype	*elavGAL4;; UAS-Q93httexon1*	*elavGAL4;;*	*w ;; UAS-Q93httexon1*
Median survival (day)	10	70	81
Curve comparison with Htt93Q : p value	/	<0.0001	<0.0001

Appendix

Table 25. Controls samples used in PRT assays.

Sample Code	Origin	Assay
C0088	ECACC	*DEFB4*
C0207	ECACC	*DEFB4*
C0849	ECACC	*DEFB4*
C0913	ECACC	*DEFB4*
C0940	ECACC	*DEFB4*
C0969	ECACC	*DEFB4*
NA19920	HapMap Panel	*SLC2A3*
NA11840	CEPH panel	*SLC2A3*
L31	Leicester panel	*SLC2A3*

Table 26. List of primers used for PCR-based assays.

Assay	Sequence (5'-3')	Size
PRT107A	Forward: AGTCCTCATTTAACTTTGGTGC Reverse: GGCTATGAAGCAATGGCCTA	155bp (reference), 158bp (test)
HPSD21	Forward: GAGGTCACTGTGATCAAAGAT Reverse: AACCTTCAGCACAGCTACTC	172bp (test), 180bp (reference)
5DELR4	Forward: AAACCAATACCCTTTCCAAG Reverse: TTCTCTTTTGTTTCAGATTCAGATG	119 to 125bp
SLC2A3	Forward: TATTGCACCTTAACCTCTCCAGC Reverse: CCTCACTTCCATACAGCTCTACG	200bp (reference), 285bp (test)
qPCR-*Glut1*	Forward: ATAATGGTTGTTATGACGCTCGTGT Reverse: GTGATGAAGATTGAGAAGATGAACA	110bp

qPCR-*RpL32**	Forward: ATCGGTTACGGATCGAACAA Reverse: ACAATCTCCTTGCGCTTCT	164bp

*from Zordan *et al.*2006

Table 27. Control fly strains.

Investigation strains	Control strains
Htt93Q/ Htt93Q*	*w;; UAS-Q93httexon1* w^{1118};; *elavGAL4*;;
Only for Htt93Q*	*elavGAL4;3M;* *elavGAL4;KK;* *elavGAL4;KK; UAS-Q93httexon1*
RNAi-*Glut1*	*w; Glut1*^{KK108683}*;* *elavGAL4; Glut1*^{KK108683}*;*
EP-*Glut1*	*w;; Glut1*^{d05758} *elavGAL4;; Glut1*^{d05758} *w;; Glut1*^{d05758}*/UAS-Q93httexon1* *elavGAL4;; Glut1* ^{d05758}
mut-*Glut1*	*w;; Glut1*^{17J} *elavGAL4;; Glut1*^{17J} *w;; Glut1*^{17J} */UAS-Q93httexon1* *elavGAL4;; Glut1*^{17J}

LITERATURE CITED

Abu Bakar S, Hollox EJ, Armour JA. 2009. Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. Proc. Natl. Acad. Sci. U. S. A. 106:853-858.

Abyzov A, Gerstein M. 2011. AGE: Defining breakpoints of genomic structural variants at single-nucleotide resolution, through optimal alignments with gap excision. Bioinformatics 27:595-603.

Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R. 2000. The new animal phylogeny: Reliability and implications. Proc. Natl. Acad. Sci. U. S. A. 97:4453-4456.

Ahuja SK, Kulkarni H, Catano G, Agan BK, Camargo JF, He W, O'Connell RJ, Marconi VC, Delmar J, Eron J. 2008. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1–infected individuals. Nat. Med. 14:413-420.

Aiken CT, Steffan JS, Guerrero CM, Khashwji H, Lukacsovich T, Simmons D, Purcell JM, Menhaji K, Zhu Y, Green K. 2009. Phosphorylation of threonine 3. J. Biol. Chem. 284:29427-29436.

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E. 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439:851-855.

Alberch J, Lopez M, Badenas C, Carrasco J, Mila M, Munoz E, Canals J. 2005. Association between BDNF Val66Met polymorphism and age at onset in Huntington disease. Neurology 65:964-965.

Al-Chalabi A, Jones A, Troakes C, King A, Al-Sarraj S, van den Berg, Leonard H. 2012. The genetics and neuropathology of amyotrophic lateral sclerosis. Acta Neuropathol. 124:339-352.

Aldhous MC, Bakar SA, Prescott NJ, Palla R, Soo K, Mansfield JC, Mathew CG, Satsangi J, Armour JAL. 2010. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. Hum Mol Genet. 19:4930-4938.

Aldred PMR, Hollox EJ, Armour JAL. 2005. Copy number polymorphism and expression level variation of the human α-defensin genes DEFA1 and DEFA3. Hum Mol Genet. 14:2045-2052.

Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. Nat Rev. Genet. 12:363-376.

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet. 41:1061-1067.

Almal SH, Padh H. 2011. Implications of gene copy-number variation in health and diseases. J. Hum. Genet. 57:6-13.

Abecasis G, Altshuler D, Auton A, Brooks L, Durbin R, Gibbs R, Hurles M, McVean G, Bentley D, Chakravarti A. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061-1073.

Amaral M, Outeiro TF, Scrutton NS, Giorgini F. 2013. The causative role and therapeutic potential of the kynurenine pathway in neurodegenerative disease. J. Mol. Med. 91(6):705-13.

Ambegaokar SS, Roy B, Jackson GR. 2010. Neurodegenerative models in Drosophila: polyglutamine disorders, Parkinson's disease, and amyotrophic lateral sclerosis. Neurobiol. Dis. 40:29-39.

Andrade MA, Bork P. 1995. HEAT repeats in the Huntington's disease protein. Nat Genet. 11:115-116.

Andresen JM, Gayán J, Cherny SS, Brocklebank D, Alkorta-Aranburu G, Addis EA, Cardon LR, Housman DE, Wexler NS. 2007. Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. J. Med. Genet. 44:44-50.

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, Squitieri F, Lin B, Kalchman MA. 1993. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet. 4:398-403.

Andrews TC, Weeks RA, Turjanski N, Gunn RN, Watkins LH, Sahakian B, Hodges JR, Rosser AE, Wood NW, Brooks DJ. 1999. Huntington's disease progression PET and clinical observations. Brain 122:2353-2363.

Arlt MF, Mulle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, Glover TW. 2009. Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. Am. J. Hum. Genet. 84:339-350.

Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ. 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res. 35:e19.

Armour JAL, Sismani C, Patsalis PC, Cross G. 2000. Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res. 28:605-609.

Arning L, Saft C, Wieczorek S, Andrich J, Kraus PH, Epplen JT. 2007. NR2A and NR2B receptor gene variations modify age at onset in Huntington disease in a sex-specific manner. Hum. Genet. 122:175-182.

Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature 431:805-810.

Augood S, Faull R, Emson P. 2004. Dopamine D1 and D2 receptor gene expression in the striatum in Huntington's disease. Ann. Neurol. 42:215-221.

Augood S, Faull R, Love D, Emson P. 1996. Reduction in enkephalin and substance P messenger RNA in the striatum of early grade Huntington's disease: A detailed cellular in situ hybridization study. Neuroscience 72:1023-1036.

Aylward EH, Liu D, Nopoulos PC, Ross CA, Pierson RK, Mills JA, Long JD, Paulsen JS. 2012. Striatal volume contributes to the prediction of onset of Huntington disease in incident cases. Biol. Psychiatry 71:822-828.

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. 2002. Recent segmental duplications in the human genome. Science 297:1003-1007.

Bence NF, Sampat RM, Kopito RR. 2001. Impairment of the ubiquitin-proteasome system by protein aggregation. Science 292:1552.

Bennardo N, Cheng A, Huang N, Stark JM. 2008. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet. 4:e1000110.

Berger EA, Murphy PM, Farber JM. 1999. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. Annu. Rev. Immunol. 17:657-700.

Besson MT, Dupont P, Fridell YW, Lievens JC. 2010. Increased energy metabolism rescues glia-induced pathology in a drosophila model of huntington's disease. Hum. Mol. Genet. 19:3372-3382.

Bhattacharya T, Stanton J, Kim E, Kunstman KJ, Phair JP, Jacobson LP, Wolinsky SM. 2009. CCL3L1 and HIV/AIDS susceptibility. Nat. Med. 15:1112-1115.

Bi W, Sapir T, Shchelochkov OA, Zhang F, Withers MA, Hunter JV, Levy T, Shinder V, Peiffer DA, Gunderson KL. 2009. Increased LIS1 expression affects human and mouse brain development. Nat Genet. 41:168-177.

Björkqvist M, Wild EJ, Thiele J, Silvestroni A, Andre R, Lahiri N, Raibon E, Lee RV, Benn CL, Soulet D. 2008. A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. J. Exp. Med. 205:1869-1877.

Blauw H, Barnes C, Van Vught P, Van Rheenen W, Verheul M, Cuppen E, Veldink J, Van Den Berg L. 2012. SMN1 gene duplications are associated with sporadic ALS. Neurology 78:776-780.

Bloomington Drosophila Stock Center at Indiana University available at: http://flystocks.bio.indiana.edu/ [Acessed August 2013].

Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas H, Hersch S, Hogarth P, Bouzou B, Jensen R. 2005. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. Proc. Natl. Acad. Sci. U. S. A. 102:11023-11028.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401-415.

Breckpot J, Thienpont B, Gewillig M, Allegaert K, Vermeesch J, Devriendt K. 2011. Differences in copy number variation between discordant monozygotic twins as a model for exploring chromosomal mosaicism in congenital heart defects. Mol. Syndromol. 2:81-87.

Brinkman R, Mezei M, Theilmann J, Almqvist E, Hayden M. 1997. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. Am. J. Hum. Genet. 60:1202.

Brown K, Heller DS, Zamudio S, Illsley NP. 2011. Glucose transporter 3 (GLUT3) protein expression in human placenta across gestation. Placenta 32:1041-1049.

Brouwers N, Van Cauwenberghe C, Engelborghs S, Lambert J, Bettens K, Le Bastard N, Pasquier F, Montoya AG, Peeters K, Mattheijssens M. 2011. Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. Mol. Psychiatry 17:223-233.

Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Ståhl T, Menzel U, Sandgren J, von Tell D, Poplawski A. 2008. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. 82:763-771.

Campesan S, Green EW, Breda C, Sathyasaikumar KV, Muchowski PJ, Schwarcz R, Kyriacou CP, Giorgini F. 2011. The kynurenine pathway modulates neurodegeneration in a Drosophila model of Huntington's disease. Curr. Biol. 21:961-966.

Cannella M, Gellera C, Maglione V, Giallonardo P, Cislaghi G, Muglia M, Quattrone A, Pierelli F, Di Donato S, Squitieri F. 2003. The gender effect in juvenile Huntington disease patients of Italian origin. Am. J. Med. Genet. B. Neuropsychiatr. Genet. 125:92-98.

Cantsilieris S, Baird PN, White SJ. 2012. Molecular methods for genotyping complex copy number polymorphisms. Genomics pii: S0888-7543(12)00206-6.

Carpenter D, Walker S, Prescott N, Schalkwijk J, Armour JA. 2011. Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders. BMC Genomics 12:418.

Carroll MC. 1998. The role of complement and complement receptors in induction and regulation of immunity. Annu. Rev. Immunol. 16:545-568.

Caviston JP, Zajac AL, Tokito M, Holzbaur EL. 2011. Huntingtin coordinates the dynein-mediated dynamic positioning of endosomes and lysosomes. Mol. Biol. Cell 22:478-492.

Cha JJ, Frey AS, Alsdorf SA, Kerner JA, Kosinski CM, Mangiarini L, Penney JB, Davies SW, Bates GP, Young AB. 1999. Altered neurotransmitter receptor expression in transgenic mouse models of Huntington's disease. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 354:981-989.

Cha JJ, Kosinski CM, Kerner JA, Alsdorf SA, Mangiarini L, Davies SW, Penney JB, Bates GP, Young AB. 1998. Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human Huntington disease gene. Proc. Natl. Acad. Sci. U.S.A. 95:6480-6485.

Cha JJ. 2007. Transcriptional signatures in Huntington's disease. Prog. Neurobiol. 83:228-248.

Chattopadhyay B, Ghosh S, Gangopadhyay PK, Das SK, Roy T, Sinha KK, Jha DK, Mukherjee SC, Chakraborty A, Singhal BS. 2003. Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India. Neurosci. Lett. 345:93-96.

Chaturvedi RK, Adhihetty P, Shukla S, Hennessy T, Calingasan N, Yang L, Starkov A, Kiaei M, Cannella M, Sassone J. 2009. Impaired PGC-1α function in muscle in Huntington's disease. Hum Mol Genet. 18:3048-3065.

Che HVB, Metzger S, Portal E, Deyle C, Riess O, Nguyen HP. 2011. Localization of sequence variations in PGC-1α influence their modifying effect in Huntington. Mol Neurodegener. 6(1):1. doi: 10.1186/1750-1326-6-1.

Chen S, Ferrone FA, Wetzel R. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. Proc. Natl. Acad. Sci. U.S.A. 99:11884-11889.

Chen JY, Wang E, Galvan L, Joshi P, Cepeda C, Levine MS. 2012. Effects of the pimelic diphenylamide histone deacetylase inhibitor HDACi 4b on the R6/2 and N171-82Q mouse models of Huntington's disease. PLoS Currents 5.

Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. 2003. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biol. 4:R25.

Cheung V, Nowak N, Jang W, Kirsch I, Zhao S, Chen XN, Furey T, Kim UJ, Kuo WL, Olivier M. 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature 409:953-958.

Chiang M, Chen H, Lee Y, Chang H, Wu Y, Soong B, Chen C, Wu Y, Liu C, Niu D. 2007. Dysregulation of C/EBPα by mutant huntingtin causes the urea cycle deficiency in Huntington's disease. Hum Mol Genet. 16:483-498.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet. 39:715-720.

Chopra V, Fox JH, Lieberman G, Dorsey K, Matson W, Waldmeier P, Housman DE, Kazantsev A, Young AB, Hersch S. 2007. A small-molecule therapeutic lead for Huntington's disease: Preclinical pharmacology and efficacy of C2-8 in the R6/2 transgenic mouse. Proc. Natl. Acad. Sci. U.S.A. 104:16685-16689.

Ciarmiello A, Giovacchini G, Orobello S, Bruselli L, Elifani F, Squitieri F. 2012. 18 F-FDG PET uptake in the pre-Huntington disease caudate affects the time-to-onset independently of CAG expansion size. Eur J Nucl Med Mol Imaging 39(6):1030-6.

Cisbani G, Cicchetti F. 2012. An in vitro perspective on the molecular mechanisms underlying mutant huntingtin protein toxicity. Cell Death Dis. 3:e382.

Clabough EB, Zeitlin SO. 2006. Deletion of the triplet repeat encoding polyglutamine within the mouse Huntington's disease gene results in subtle behavioral/motor phenotypes in vivo and elevated levels of ATP with cellular senescence in vitro. Hum Mol Genet. 15:607-623.

Clustal Omega Multiple Sequence Alignment available at http://www.ebi.ac.uk/Tools/msa/clustalo/ [Acessed August 2013].

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P *et al*.„ (x co-authors. 2010. Origins and functional impact of copy number variation in the human genome. Nature 464:704-712.

Cook Jr EH, Scherer SW. 2008. Copy-number variations associated with neuropsychiatric conditions. Nature 455:919-923.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V. 2011. A copy number variation morbidity map of developmental delay. Nat Genet. 43:838-846.

Cooper GM, Nickerson DA, Eichler EE. 2007. Mutational and selective effects on copy-number variants in the human genome. Nat Genet. 39:S22-S29.

Cornett J, Cao F, Wang C, Ross CA, Bates GP, Li S, Li X. 2005. Polyglutamine expansion of huntingtin impairs its nuclear export. Nat Genet. 37:198-204.

Dalrymple A, Wild EJ, Joubert R, Sathasivam K, Björkqvist M, Petersén Å, Jackson GS, Isaacs JD, Kristiansen M, Bates GP. 2007. Proteomic profiling of plasma in Huntington's disease reveals neuroinflammatory activation and biomarker candidates. J Proteome Res 6:2833-2840.

Database of Genomic Variants available at: http://dgvbeta.tcag.ca/dgv/app/home?ref=NCBI36/hg18 [Acessed August 2013].

Davies SW, Turmaine M, Cozens BA, DiFiglia M, Sharp AH, Ross CA, Scherzinger E, Wanker EE, Mangiarini L, Bates GP. 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. Cell 90:537-548.

Dermitzakis ET, Stranger BE. 2006. Genetic variation in human gene expression. Mammalian Genome 17:503-508.

Dhaenens CM, Burnouf S, Simonin C, Van Brussel E, Duhamel A, Defebvre L, Duru C, Vuillaume I, Cazeneuve C, Charles P *et al*.(x co-authors. 2009. A genetic variation in the ADORA2A gene modifies age at onset in Huntington's disease. Neurobiol. Dis. 35:474-476.

Di Maria E, Marasco A, Tartari M, Ciotti P, Abbruzzese G, Novelli G, Bellone E, Cattaneo E, Mandich P. 2006. No evidence of association between BDNF gene variants and age-at-onset of Huntington's disease. Neurobiol. Dis. 24:274-279.

Diekmann H, Anichtchik O, Fleming A, Futter M, Goldsmith P, Roach A, Rubinsztein DC. 2009. Decreased BDNF levels are a major contributor to the embryonic phenotype of huntingtin knockdown zebrafish. J. Neurosci.29:1343-1349.

Dietzl G, Chen D, Schnorrer F, Su K, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-156.

DiFiglia M, Sapp E, Chase K, Schwarz C, Meloni A, Young C, Martin E, Vonsattel J, Carraway R, Reeves SA. 1995. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. Neuron 14:1075-1081.

DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel J, Aronin N. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 277:1990-1993.

Dorsey ER. 2012. Characterization of a large group of individuals with Huntington disease and their relatives enrolled in the COHORT study. PloS One 7:e29522.

Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S. 2006. A network of protein interactions determines polyglutamine toxicity. Proc. Natl. Acad. Sci. U.S.A. 103:11051-11056.

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M. 1993. Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet. 4:387-392.

Duyao MP, Auerbach AB, Ryan A, Persichetti F. 1995. Inactivation of the mouse Huntington's disease gene homolog Hdh. Science 269(5222):407-10.

Eichler EE. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet.17:661-669.

Engelender S, Sharp AH, Colomer V, Tokito MK, Lanahan A, Worley P, Holzbaur EL, Ross CA. 1997. Huntingtin-associated protein 1 (HAP1) interacts with the p150Glued subunit of dynactin. Hum Mol Genet. 6:2205-2212.

Escher SA, Rasmuson-Lestander A. 1999. The Drosophila glucose transporter gene: CDNA sequence, phylogenetic comparisons, analysis of functional sites and secondary structures. Hereditas. 130(2):95-103.

Espejel S, Franco S, Rodriguez-Perales S, Bouffler SD, Cigudosa JC, Blasco MA. 2002. Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres. Embo J. 21:2207-2219.

Falush D. 2009. Haplotype background, repeat length evolution, and Huntington's disease. Am. J. Hum. Genet. 85:939.

Fan MM, Raymond LA. 2007. N-methyl-d-aspartate (NMDA) receptor function and excitotoxicity in Huntington's disease. Prog. Neurobiol. 81:272-293.

Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, Heward JM, Gough SC, de Smith A, Blakemore AI. 2007. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet. 39:721-723.

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, Maraganore D, Gwinn-Hardy K, Wszolek Z, Dickson D. 2004. Comparison of kindreds with Parkinsonism and α-synuclein genomic multiplications. Ann. Neurol. 55:174-179.

Feigin A, Leenders KL, Moeller JR, Missimer J, Kuenig G, Spetsieris P, Antonini A, Eidelberg D. 2001. Metabolic network abnormalities in early Huntington’s disease: An [18F] FDG PET study. J. Nucl. Med. 42:1591-1595.

Ferrante RJ, Kubilus JK, Lee J, Ryu H, Beesen A, Zucker B, Smith K, Kowall NW, Ratan RR, Luthi-Carter R. 2003. Histone deacetylase inhibition by sodium butyrate chemotherapy ameliorates the neurodegenerative phenotype in Huntington's disease mice. J. Neurosci.23:9418-9427.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nature Reviews.Genetics 7:85-97.

Field S, Howson J, Maier L, Walker S, Walker N, Smyth D, Armour J, Clayton D, Todd J. 2009. Experimental aspects of copy number variant assays at CCL3L1. Nat. Med. 15(10):1115-1117.

Fly base genome browser available at : http://flybase.org/cgi-bin/gbrowse2/dmel/ [Acessed August 2013].

FlyAtlas: the *Drosophila* gene expression atlas available at: http://www.flyatlas.org/ [Acessed August 2013].

Fode P, Jespersgaard C, Hardwick RJ, Bogle H, Theisen M, Dodoo D, Lenicek M, Vitek L, Vieira A, Freitas J. 2011. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PLOS One 6:e16768.

Frank S, Ondo W, Fahn S, Hunter C, Oakes D, Plumb S, Marshall F, Shoulson I, Eberly S, Walker F. 2008. A study of chorea after tetrabenazine withdrawal in patients with Huntington disease. Clin. Neuropharmacol. 31:127-133.

Fuchs J, Nilsson C, Kachergus J, Munz M, Larsson EM, Schüle B, Langston J, Middleton F, Ross O, Hulihan M. 2007. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology 68:916-922.

Fusco FR, Chen Q, Lamoreaux WJ, Figueredo-Cardenas G, Jiao Y, Coffman JA, Surmeier DJ, Honig MG, Carlock LR, Reiner A. 1999. Cellular localization of huntingtin in striatal and cortical neurons in rats: lack of correlation with neuronal vulnerability in Huntington's disease. J. Neurosci.19:1189-1202.

Gafni J, Ellerby LM. 2002. Calpain activation in Huntington's disease. J. Neurosci.22:4842-4849.

Gafni J, Hermel E, Young JE, Wellington CL, Hayden MR, Ellerby LM. 2004. Inhibition of calpain cleavage of huntingtin reduces toxicity. J. Biol. Chem. 279:20211-20220.

Ganz T. 2003. Defensins: antimicrobial peptides of innate immunity. Nat Rev. Immunol. 3:710-720.

Gaucher C, Mazurier C. 1995. Characterization of factor VIII gene inversions using a non-radioactive detection method: A survey of 102 unrelated haemophilia A families from northern France. Nouv. Rev. Fr. Hematol. 37:131-131.

Gauthier LR, Charrin BC, Borrell-Pagès M, Dompierre JP, Rangone H, Cordelières FP, De Mey J, MacDonald ME, Leßmann V, Humbert S. 2004. Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. Cell 118:127-138.

Georgiou-Karistianis N, Scahill R, Tabrizi SJ, Squitieri F, Aylward E. 2013. Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. Neurosci Biobehav Rev.37(3):480-90.

Ghani M, Pinto D, Lee JH, Grinberg Y, Sato C, Moreno D, Scherer SW, Mayeux R, George-Hyslop PS, Rogaeva E. 2012. Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean Hispanics. G3 2:71-78.

Gil JM, Rego AC. 2009. The R6 lines of transgenic mice: a model for screening new therapies for Huntington's disease. Brain Res. Rev. 59:410-431.

Giorgini F, Guidetti P, Nguyen QV, Bennett SC, Muchowski PJ. 2005. A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. Nat Genet. 37:526-531.

Giorgini F, Möller T, Kwan W, Zwilling D, Wacker JL, Hong S, Tsai LCL, Cheah CS, Schwarcz R, Guidetti P. 2008. Histone deacetylase inhibition modulates kynurenine pathway activation in yeast, microglia, and mice expressing a mutant huntingtin fragment. J. Biol. Chem. 283:7390-7400.

Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C. 2010. A recurrent 16p12. 1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet. 42:203-209.

Goldberg Y, Nicholson D, Rasper D, Kalchman M, Koide H, Graham R, Bromm M, Kazemi-Esfarjani P, Thornberry N, Vaillancourt J. 1996. Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. Nat Genet. 13:442-449.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434.

Graham RK, Deng Y, Slow EJ, Haigh B, Bissada N, Lu G, Pearson J, Shehadeh J, Bertram L, Murphy Z. 2006. Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. Cell 125:1179-1191.

Green E, Giorgini F. 2012. Choosing and using Drosophila models to characterize modifiers of Huntington's disease. Biochem. Soc. Trans. 40:739.

Green EW, Campesan S, Breda C, Sathyasaikumar KV, Muchowski PJ, Schwarcz R, Kyriacou CP, Giorgini F. 2012. Drosophila eye color mutants as therapeutic tools for Huntington disease. Fly 6:117-120.

Gunawardena S, Her L, Brusch RG, Laymon RA, Niesman IR, Gordesky-Gold B, Sintasath L, Bonini NM, Goldstein LS. 2003. Disruption of axonal transport by loss of huntingtin or expression of pathogenic PolyQ proteins in Drosophila. Neuron 40:25-40.

Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SAAC, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD. 2008. Distribution and functional impact of DNA copy number variation in the rat. Nat Genet. 40:538-545.

Gusella JF, MacDonald ME. 2009. Huntington's disease: the case for genetic modifiers. Genome Med 1:80.

Hackam AS, Singaraja R, Wellington CL, Metzler M, McCutcheon K, Zhang T, Kalchman M, Hayden MR. 1998. The influence of huntingtin protein size on nuclear localization and cellular toxicity. J. Cell Biol. 141:1097-1105.

Haldimann P, Muriset M, Vígh L, Goloubinoff P. 2011. The novel hydroxylamine derivative NG-094 suppresses polyglutamine protein toxicity in caenorhabditis elegans. J. Biol. Chem. 286:18784-18794.

Hardwick RJ, Machado LR, Zuccherato LW, Antolinos S, Xue Y, Shawa N, Gilman RH, Cabrera L, Berg DE, Tyler-Smith C. 2011. A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. Hum. Mutat. 32:743-750.

Harper PS. 1992. The epidemiology of Huntington's disease. Hum. Genet. 89:365-376.

Hastings P, Ira G, Lupski JR. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genetics 5:e1000327.

Hastings P, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. Nat Rev. Genet. 10:551-564.

Heather West Greenlee M, Uemura E, Carpenter SL, Doyle RT, Buss JE. 2003. Glucose uptake in PC12 cells: GLUT3 vesicle trafficking and fusion as revealed with a novel GLUT3-GFP fusion protein. J. Neurosci. Res. 73:518-525.

Heinzen EL, Need AC, Hayden KM, Chiba-Falek O, Roses AD, Strittmatter WJ, Burke JR, Hulette CM, Welsh-Bohmer KA, Goldstein DB. 2010. Genome-wide scan of copy number variation in late-ons*et al*zheimer's disease. J Alzheimer's Dis. 19:69-77.

Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, De Kovel C, Baker C, Von Spiczak S. 2009. 15q13. 3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 41:160-162.

Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M, Kaessmann H, Reymond A. 2009. Segmental copy number variation shapes tissue transcriptomes. Nat Genet. 41:424-429.

Hermel E, Gafni J, Propp S, Leavitt B, Wellington C, Young J, Hackam A, Logvinova A, Peel A, Chen S. 2004. Specific caspase interactions and amplification are involved in selective neuronal vulnerability in Huntington's disease. Cell Death Differ. 11:424-438.

Higuchi R, Fockler C, Dollinger G, Watson R. 1993. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. Biotechnology 11:1026-1030.

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D. 2006. Regional and cellular gene expression changes in human

Huntington's disease brain. Hum Mol Genet. 15:965-977.

Hollox E, Detering J, Dehnugara T. 2009. An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. Hum. Mutat. 30(3):477-484.

Hollox EJ, Armour JAL, Barber JCK. 2003. Extensive normal copy number variation of a β-defensin antimicrobial-gene cluster. Am. J. Hum. Genet.73:591-600.

Hollox EJ. 2010. Editorial: β-defensins and Crohn's disease: confusion from counting copies. Am. J. Gastroenterol. 105:360-362.

Hollox EJ. 2012. The challenges of studying complex and dynamic regions of the human genome. Methods Mol. Biol. 838:187-207.

Holmberg CI, Staniszewski KE, Mensah KN, Matouschek A, Morimoto RI. 2004. Inefficient degradation of truncated polyglutamine proteins by the proteasome. Embo j. 23:4307-4318.

Hsiao, H., Y. Chen, H. Chen, P. Tu, and Y. Chern. 2013. A critical role of astrocyte-mediated nuclear factor-κB-dependent inflammation in Huntington's disease. Hum Mol Genet. 22(9):1826-42.

Huang J, Wei W, Zhang J, Liu G, Bignell GR, Stratton MR, Futreal PA, Wooster R, Jones KW, Shapero MH. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum. Genomics 1:287-299.

Hurlbert MS, Zhou W, Wasmeier C, Kaddis F, Hutton J, Freed C. 1999. Mice transgenic for an expanded CAG repeat in the Huntington's disease gene develop diabetes. Diabetes 48:649-651.

Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. Trends Genet. 24:238-245.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nat Genet. 36:949-951.

Ibanez P, Bonnet A, Debarges B, Lohmann E, Tison F, Pollak P, Agid Y, Dürr A, Brice A. 2004. Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease. The Lancet 364:1169-1171.

Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, Rosenbarker L, Bird TD, Hodes M, Shaffer LG, Lupski JR. 2002. Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. Am. J. Hum. Genet. 71:838-853.

International Conference on Harmonisation Guidelines available at : http://www.ich.org/ [Acessed August 2013].

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437(7063):1299-1320.

Iskow RC, Gokcumen O, Lee C. 2012. Exploring the role of copy number variants in human adaptation. Trends Genet. 28(6):245-57.

Jackson GR, Salecker I, Dong X, Yao X, Arnheim N, Faber PW, MacDonald ME, Zipursky SL. 1998. Polyglutamine-expanded human huntingtin transgenes induce degeneration of Drosophila photoreceptor neurons. Neuron 21:633-642.

Jacobs PA, Baikie AG, Court Brown WM, Strong JA. 1959. The somatic chromosomes in mongolism. The Lancet 1:710.

Jansen PA, Rodijk-Olthuis D, Hollox EJ, Kamsteeg M, Tjabringa GS, de Jongh GJ, van Vlijmen-Willems IM, Bergboer JG, van Rossum MM, de Jong EM. 2009. β-Defensin-2 protein is a serum biomarker for disease activity in psoriasis and reaches biologically relevant concentrations in lesional skin. PLoS One 4:e4725.

Jaradat S, Hoder-Przyrembel C, Cubillos S, Krieg N, Lehmann K, Piehler S, Sigusch B, Norgauer J. 2013. Beta-defensin-2 genomic copy number variation and chronic periodontitis. J. Dent. Res. :0022034513504217.

Janssens W, Nuytten H, Dupont LJ, Van Eldere J, Vermeire S, Lambrechts D, Nackaerts K, Decramer M, Cassiman J, Cuppens H. 2010. Genomic copy number determines functional expression of β-defensin 2 in airway epithelial cells and associates with chronic obstructive

pulmonary disease. American Journal of Respiratory and Critical Care Medicine 182:163-169.

Jia H, Pallos J, Jacques V, Lau A, Tang B, Cooper A, Syed A, Purcell J, Chen Y, Sharma S. 2012. Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington's disease. Neurobiol. Dis. 46:351-361.

Jenkins BG, Koroshetz WJ, Beal MF, Rosen BR. 1993. Evidence for irnnairment of energy metabofism in vivo in Huntington's disease using localized 1H NMR spectroscopy. Neurology 43:2689-2689.

Josefsen K, Nielsen M, Jørgensen K, Bock T, Nørremølle A, Sørensen SA, Naver B, Hasholt L. 2007. Impaired glucose tolerance in the R6/1 transgenic mouse model of Huntington's disease. J. Neuroendocrinol. 20:165-172.

Ju T, Lin Y, Chern Y. 2012. Energy dysfunction in Huntington's disease: Insights from PGC-1α, AMPK, and CKB. Cell. Mol. Life Sci. 69(24):4107-20

Kabashi E, Brustein E, Champagne N, Drapeau P. 2011. Zebrafish models for the functional genomics of neurogenetic disorders. Biochim. Biophys. Acta. 1812:335-345.

Kalchman MA, Graham RK, Xia G, Koide HB, Hodgson JG, Graham KC, Goldberg YP, Gietz RD, Pickart CM, Hayden MR. 1996. Huntingtin is ubiquitinated and interacts with a specific ubiquitin-conjugating enzyme. J. Biol. Chem. 271:19385-19394.

Kaltenbach LS, Romero E, Becklin RR, Chettier R, Bell R, Phansalkar A, Strand A, Torcassi C, Savage J, Hurlburt A. 2007. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. PLoS Genetics 3:e82.

Karadima G, Dimovasili C, Koutsis G, Vassilopoulos D, Panas M. 2012. Age at onset in Huntington's disease: Replication study on the association of HAP1. Parkinsonism Relat. Disord. 18(9):1027-8.

Karlovich CA, John RM, Ramirez L, Stainier DY, Myers RM. 1998. Characterization of the Huntington's disease (HD) gene homolog in the zebrafish Danio rerio. Gene 217:117-125.

Kaye JA, Finkbeiner S. 2013. Modeling Huntington's disease with induced pluripotent stem cells. Mol Cell Neurosci. 56C:50-64.

Kazemi-Esfarjani P, Benzer S. 2000. Genetic suppression of polyglutamine toxicity in Drosophila. Science 287:1837-1840.

Kennedy L, Evans E, Chen C, Craven L, Detloff PJ, Ennis M, Shelbourne PF. 2003. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet. 12:3359-3367.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F *et al*.,, (x co-authors. 2008. Mapping and sequencing of structural variation from eight human genomes. Nature 453:56-64.

Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. 2007. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genetics 3:e63.

Kidd JM, Sampas N, Antonacci F, Graves T, Fulton R, Hayden HS, Alkan C, Malig M, Ventura M, Giannuzzi G. 2010. Characterization of missing human genome sequences and copy-number polymorphic insertions. Nat. Methods 7:365-371.

Kim YJ, Yi Y, Sapp E, Wang Y, Cuiffo B, Kegel KB, Qin Z, Aronin N, DiFiglia M. 2001. Caspase 3-cleaved N-terminal fragments of wild-type and mutant huntingtin are present in normal and Huntington's disease brains, associate with membranes, and undergo calpain-dependent proteolysis. Proc. Natl. Acad. Sci. U.S.A. 98:12784-12789.

Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. 2009. Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophr. Bull. 35:851-854.

Kishikawa S, Li J, Gillis T, Hakky MM, Warby S, Hayden M, MacDonald ME, Myers RH, Gusella JF. 2006. Brain-derived neurotrophic factor does not influence age at neurologic onset of Huntington’s disease. Neurobiol. Dis. 24:280-285.

Kita H, Carmichael J, Swartz J, Muro S, Wyttenbach A, Matsubara K, Rubinsztein DC, Kato K. 2002. Modulation of polyglutamine-induced cell death by genes identified by expression profiling. Hum Mol Genet. 11:2279-2287.

Kitamura E, Blow JJ, Tanaka TU. 2006. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. Cell 125:1297.

Korbel JO, Abyzov A, Mu XJ, Carriero N, Cayting P, Zhang Z, Snyder M, Gerstein MB. 2009. PEMer: A computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genome Biol. 10:R23.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L. 2007. Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420-426.

Kremer B, Almqvist E, Theilmann J, Spence N, Telenius H, Goldberg Y, Hayden M. 1995. Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am. J. Hum. Genet. 57:343.

Kuhl DE, Phelps ME, Markham CH, Metter EJ, Riege WH, Winter J. 2004. Cerebral metabolism and atrophy in Huntington's disease determined by 18FDG and computed tomographic scan. Ann. Neurol. 12:425-434.

Kurotaki N, Shen JJ, Touyama M, Kondoh T, Visser R, Ozaki T, Nishimoto J, Shiihara T, Uetake K, Makita Y. 2005. Phenotypic consequences of genetic variation at hemizygous alleles: Sotos syndrome is a contiguous gene syndrome incorporating coagulation factor twelve (FXII) deficiency. Genet. Med. 7:479-483.

Landles C, Sathasivam K, Weiss A, Woodman B, Moffitt H, Finkbeiner S, Sun B, Gafni J, Ellerby LM, Trottier Y. 2010. Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. J. Biol. Chem. 285:8808-8823.

Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden M. 2004. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin. Genet. 65:267-277.

Langbehn DR, Hayden MR, Paulsen JS. 2010. CAG-repeat length and the age of onset in Huntington disease (HD): A review and validation study of statistical approaches. Am J Med Genet B Neuropsychiatr Genet. 153:397-408.

Lange J, Skaletsky H, van Daalen SK, Embry SL, Korver CM, Brown LG, Oates RD, Silber S, Repping S, Page DC. 2009. Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell 138:855-869.

Leavitt BR, Raamsdonk JM, Shehadeh J, Fernandes H, Murphy Z, Graham RK, Wellington CL, Hayden MR. 2006. Wild-type huntingtin protects neurons from excitotoxicity. J. Neurochem. 96:1121-1129.

Lee J, Gillis T, Mysore JS, Ramos EM, Myers RH, Hayden MR, Morrison PJ, Nance M, Ross CA, Margolis RL. 2012c. Common SNP-based haplotype analysis of the 4p16. 3 Huntington disease gene region. Am. J. Hum. Genet. 90(3):434-44

Lee J, Lee J, Ramos EM, Gillis T, Mysore JS, Kishikawa S, Hadzi T, Hendricks AE, Hayden MR, Morrison PJ. 2012a. TAA repeat variation in the GRIK2 gene does not influence age at onset in Huntington's disease. Biochem. Biophys. Res. Commun. 424(3):404-8

Lee J, Ramos E, Lee J, Gillis T, Mysore J, Hayden M, Warby S, Morrison P, Nance M, Ross C. 2012b. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. Neurology 78:690-695.

Lee JA, Carvalho C, Lupski JR. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131:1235-1247.

Lee WM, Yoshihara M, Littleton JT. 2004. Cytoplasmic aggregates trap polyglutamine-containing proteins and block axonal transport in a Drosophila model of Huntington's disease. Proc. Natl. Acad. Sci. U. S. A. 101:3224-3229.

Lejeune F, Mesrob L, Parmentier F, Bicep C, Vazquez-Manrique RP, Parker JA, Vert J, Tourette C, Neri C. 2012. Large-scale functional RNAi screen in C. elegans identifies genes that regulate the dysfunction of mutant polyglutamine neurons. BMC Genomics 13:91.

Li H, Li S, Yu Z, Shelbourne P, Li X. 2001. Huntingtin aggregate-associated axonal degeneration is an early pathological event in Huntington's disease mice. J. Neurosci.21:8473-8481.

Li S, Li X. 1998. Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. Hum Mol Genet. 7:777-782.

Li S, Schilling G, Young W, Li X, Margolis R, Stine O, Wagster M, Abbott M, Franz M, Ranen N. 1993. Huntington's disease gene (IT15) is widely expressed in human and rat tissues. Neuron 11:985-993.

Li X, Li S, Sharp AH, Nucifora FC, Schilling G, Lanahan A, Worley P, Snyder SH, Ross CA. 1995. A huntingtin-associated protein enriched in brain with implications for pathology. Nature 378(6555):398-402.

Li X, Sapp E, Chase K, Comer-Tierney LA, Masso N, Alexander J, Reeves P, Kegel KB, Valencia A, Esteves M. 2009a. Disruption of Rab11 activity in a knock-in mouse model of Huntington's disease. Neurobiol. Dis. 36:374-383.

Li X, Standley C, Sapp E, Valencia A, Qin Z, Kegel KB, Yoder J, Comer-Tierney LA, Esteves M, Chase K. 2009b. Mutant huntingtin impairs vesicle formation from recycling endosomes by interfering with Rab11 activity. Mol. Cell. Biol. 29:6106-6116.

Li X, Valencia A, McClory H, Sapp E, Kegel KB, DiFiglia M. 2012. Deficient Rab11 activity underlies glucose hypometabolism in primary neurons of Huntington's disease mice. Biochem. Biophys. Res. Commun. 421(4):727-30.

Lieber MR, Ma Y, Pannicke U, Schwarz K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev. Mol. Cell. Biol. 4(9):712-720.

Lieber MR. 2008. The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. 283:1-5.

Lifton RP, Dluhy RG, Powers M, Rich GM, Cook S, Ulick S, Lalouel JM. 1992. A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. Nature 355(6357):262-265.

Lin B, Rommens JM, Graham RK, Kalchman M, MacDonald H, Nasir J, Delaney A, Goldberg YP, Hayden MR. 1993. Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. Hum Mol Genet. 2:1541-1545.

Liu P, Carvalho C, Hastings P, Lupski JR. 2012. Mechanisms for recurrent and complex human genomic rearrangements. Curr. Opin. Genet. Dev. 22(3):211-20.

Liu P, Lacaria M, Zhang F, Withers M, Hastings P, Lupski JR. 2011. Frequency of nonallelic homologous recombination is correlated with length of homology: Evidence that ectopic synapsis precedes ectopic crossing-over. Am. J. Hum. Genet.89:580-588.

Lu B, Palacino J. 2013. A novel human embryonic stem cell-derived Huntington's disease neuronal model exhibits mutant huntingtin (mHTT) aggregates and soluble mHTT-dependent neurodegeneration. FASEB J. 27(5):1820-9.

Lucotte G, Turpin J, Riess O, Epplen J, Siedlaczk I, Loirat F, Hazout S. 1995. Confidence intervals for predicted age of onset, given the size of (CAG) n repeat, in Huntington's disease. Hum. Genet. 95:231-232.

Lumsden AL, Henshall TL, Dayan S, Lardelli MT, Richards RI. 2007. Huntingtin-deficient zebrafish exhibit defects in iron utilization and development. Hum Mol Genet. 16:1905-1920.

Lunkes A, Lindenberg KS, Ben-Haïem L, Weber C, Devys D, Landwehrmeyer GB, Mandel J, Trottier Y. 2002. Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions. Mol. Cell 10:259-269.

Luo Y, Hermetz KE, Jackson JM, Mulle JG, Dodd A, Tsuchiya KD, Ballif BC, Shaffer LG, Cody JD, Ledbetter DH. 2011b. Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. Hum Mol Genet. 20:3769-3778.

Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, Barker DF, Killian JM, Garcia CA. 1991b. DNA duplication associated with charcot-marie-tooth disease type 1A. Cell 66:219-232.

Lupski JR, Stankiewicz PT. 2006. Genomic disorders: The genomic basis of disease. Humana Press.

Lupski JR. 2007. Genomic rearrangements and sporadic disease. Nat Genet. 39:S43-S47.

Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, Frey AS, Spektor BS, Penney EB, Schilling G. 2000. Decreased expression of striatal signaling genes in a mouse model of Huntington’s disease. Hum Mol Genet. 9:1259-1271.

Luthi-Carter R, Strand AD, Hanson SA, Kooperberg C, Schilling G, La Spada AR, Merry DE, Young AB, Ross CA, Borchelt DR. 2002. Polyglutamine and transcription: Gene expression changes shared by DRPLA and Huntington's disease mouse models reveal context-independent effects. Hum Mol Genet. 11:1927-1937.

Luthi-Carter R, Taylor DM, Pallos J, Lambert E, Amore A, Parker A, Moffitt H, Smith DL, Runne H, Gokce O. 2010. SIRT2 inhibition achieves neuroprotection by decreasing sterol biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 107:7927-7932.

Lv Y, He S, Zhang Z, Li Y, Hu D, Zhu K, Cheng H, Zhou F, Chen G, Zheng X. 2011. Confirmation of C4 gene copy number variation and the association with systemic lupus erythematosus in Chinese Han population. Rheumatol. Int. 32(10):3047-53.

MacDonald M, Vonsattel J, Shrinidhi J, Couropmitree N, Cupples L, Bird E, Gusella J, Myers R. 1999. Evidence for the GluR6 gene associated with younger onset age of Huntington's disease. Neurology 53:1330-1330.

MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971-983.

Mackay IR, Rosen FS, Walport MJ. 2001. Complement. N Engl J Med. 344:1058-1066.

Mai M, Akkad AD, Wieczorek S, Saft C, Andrich J, Kraus PH, Epplen JT, Arning L. 2006. No association between polymorphisms in the BDNF gene and age at onset in Huntington disease. BMC Med Genet 7:79.

Mamtani M, Anaya J, He W, Ahuja S. 2009. Association of copy number variation in the FCGR3B gene with risk of autoimmune diseases. Genes Immun. 11:155-160.

Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, Lawton M, Trottier Y, Lehrach H, Davies SW. 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell 87:493-506.

Marin EC, Jefferis GS, Komiyama T, Zhu H, Luo L. 2002. Representation of the glomerular olfactory map in the Drosophila brain. Cell 109:243-255.

Marques RB, Thabet MM, White SJ, Houwing-Duistermaat JJ, Bakker AM, Hendriks G, Zhernakova A, Huizinga TW, van der Helm-van, Annette H, Toes RE. 2010. Genetic variation of the fc gamma receptor 3B gene and association with rheumatoid arthritis. PloS One 5:e13173.

Marsh JL, Pallos J, Thompson LM. 2003. Fly models of Huntington's disease. Hum Mol Genet. 12 Spec No 2:R187-93.

Marsh JL, Walker H, Theisen H, Zhu YZ, Fielder T, Purcell J, Thompson LM. 2000. Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in Drosophila. Hum Mol Genet. 9:13-25.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y. 2008. Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet.82:477-488.

Martindale D, Hackam A, Wieczorek A, Ellerby L, Wellington C, McCutcheon K, Singaraja R, Kazemi-Esfarjani P, Devon R, Kim SU. 1998. Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. Nat Genet. 18:150-154.

Mason RP, Giorgini F. 2011. Modeling Huntington disease in yeast: Perspectives and future directions. Prion 5:269-276.

McGuire JR, Rong J, Li S, Li X. 2006. Interaction of huntingtin-associated protein-1 with kinesin light chain. J. Biol. Chem. 281:3552-3559.

McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, Jones PB, McLean L, O'Donnell JL, Pokorny V. 2008. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Ann. Rheum. Dis. 67:409-413.

McKinney C, Merriman TR. 2012. Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes. Hum Mol Genet. 21:2370-2376.

McLellan D, Chalmers R, Johnson R. 1974. A double-blind trial of tetrabenazine, thiopropazate, and placebo in patients with chorea. The Lancet 303:104-107.

McNeil SM, Novelletto A, Srinidhi J, Barnes G, Kornbluth I, Altherr MR, Wasmuth JJ, Gusella JF, MacDonald ME, Myers RH. 1997. Reduced penetrance of the Huntington's disease

mutation. Hum Mol Genet. 6:775-779.

Medvedev P, Stanciu M, Brudno M. 2009. Computational methods for discovering structural variation with next-generation sequencing. Nature Methods 6:S13-S20.

Mefford HC, Eichler EE. 2009. Duplication hotspots, rare genomic disorders, and common disease. Curr. Opin. Genet. Dev. 19:196-204.

Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla JA, Baralle D. 2008. Recurrent rearrangements of chromosome 1q21. 1 and variable pediatric phenotypes. N Engl J Med 359:1685-1699.

Meriin AB, Zhang X, He X, Newnam GP, Chernoff YO, Sherman MY. 2002. Huntingtin toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1. J. Cell Biol. 157:997-1004.

Meßmer K, Reynolds GP. 1998a. Increased peripheral benzodiazepine binding sites in the brain of patients with Huntington's disease. Neurosci Lett. 241:53-56.

Menalled LB. 2005. Knock-in mouse models of Huntington's disease. Neurorx 2:465-470.

Metzger S, Bauer P, Tomiuk J, Laccone F, DiDonato S, Gellera C, Mariotti C, Lange HW, Weirich-Schwaiger H, Wenning GK. 2006a. Genetic analysis of candidate genes modifying the age-at-onset in Huntington’s disease. Hum Genet. 120:285-292.

Metzger S, Bauer P, Tomiuk J, Laccone F, Didonato S, Gellera C, Soliveri P, Lange HW, Weirich-Schwaiger H, Wenning GK *et al*. (x co-authors. 2006b. The S18Y polymorphism in the UCHL1 gene is a genetic modifier in Huntington's disease. Neurogenetics 7:27-30.

Metzger S, Rong J, Nguyen HP, Cape A, Tomiuk J, Soehn AS, Propping P, Freudenberg-Hua Y, Freudenberg J, Tong L *et al*.(x co-authors. 2008. Huntingtin-associated protein-1 is a modifier of the age-at-onset of Huntington's disease. Hum Mol Genet. 17:1137-1146.

Metzker ML. 2009. Sequencing technologies—the next generation. Nat Rev Genet. 11:31-46.

Miller DW, Hague SM, Clarimo Nature Reviews Genetics J, Baptista M, Gwinn-Hardy K, Cookson MR, Singleton AB. 2004. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. Neurology 62:1835-1838.

Miller JP, Holcomb J, Al-Ramahi I, de Haro M, Gafni J, Zhang N, Kim E, Sanhueza M, Torcassi C, Kwak S. 2010. Matrix metalloproteinases are modifiers of huntingtin proteolysis and toxicity in Huntington's disease. Neuron 67:199-212.

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59-65.

Mitra S, Tsvetkov AS, Finkbeiner S. 2009. Single neuron ubiquitin-proteasome dynamics accompanying inclusion body formation in Huntington disease. J. Biol. Chem. 284:4398-4403.

Morris DL, Roberts AL, Witherden AS, Tarzi R, Barros P, Whittaker JC, Cook TH, Aitman TJ, Vyse TJ. 2010. Evidence for both copy number and allelic (NA1/NA2) risk at the FCGR3B locus in systemic lupus erythematosus. Eur J Hum Genet. 18:1027-1031.

Morton A, Faull R, Edwardson J. 2001. Abnormalities in the synaptic vesicle fusion machinery in Huntington’s disease. Brain Res. Bull. 56:111-117.

Moura KC, Junior MC, de Rosso, Ana Lucia Zuma, Nicaretta DH, Pereira JS, Jose Silva D, Santos-Reboucas CB, Pimentel MMG. 2012. Exon dosage variations in Brazilian patients with Parkinson's disease: analysis of SNCA, PARKIN, PINK1 and DJ-1 genes. Dis. Markers 32(3):173-178.

Myers R, Leavitt J, Farrer L, Jagadeesh J, McFarlane H, Mastromauro C, Mark R, Gusella J. 1989. Homozygote for Huntington disease. Am. J. Hum. Genet. 45:615.

Nagai Y, Fujikake N, Ohno K, Higashiyama H, Popiel HA, Rahadian J, Yamaguchi M, Strittmatter WJ, Burke JR, Toda T. 2003. Prevention of polyglutamine oligomerization and neurodegeneration by the peptide inhibitor QBP1 in Drosophila. Hum Mol Genet. 12:1253-1259.

Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, Borowski A, Marth JD, Phillips AG, Hayden MR. 1995. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. Cell

81:811-823.

Nathans J, Piantanida TP, Eddy RL, Shows TB, Hogness DS. 1986. Molecular genetics of inherited variation in human color vision. Science 232(4747):203-210.

Naze P, Vuillaume I, Destee A, Pasquier F, Sablonniere B. 2002. Mutation analysis and association studies of the ubiquitin carboxy-terminal hydrolase L1 gene in Huntington's disease. Neurosci. Lett. 328:1-4.

Nguyen D, Webber C, Hehir-Kwa J, Pfundt R, Veltman J, Ponting CP. 2008. Reduced purifying selection prevails over positive selection in human copy number variant evolution. Genome Res. 18:1711-1723.

Nguyen DQ, Webber C, Ponting CP. 2006. Bias of selection on human copy-number variants. PLoS Genetics 2:e20.

Niederer HA, Clatworthy MR, Willcocks LC, Smith KG. 2010. FcγRIIB, FcγRIIIB, and systemic lupus erythematosus. Ann. N. Y. Acad. Sci. 1183:69-88.

Nossent JC, Rischmueller M, Lester S. 2012. Low copy number of the fc-γ receptor 3B gene FCGR3B is a risk factor for primary Sjögren's syndrome. J. Rheumatol. 39:2142-2147.

Notini A, Craig J, White S. 2008. Copy number variation and mosaicism. Cytogenet Genome Res. 123:270-277.

Novak M, Tabrizi SJ. 2011. Huntington's disease: clinical presentation and treatment. Int. Rev. Neurobiol. 98:297-323.

Olsson LM, Holmdahl R. 2012. Copy number variation in autoimmunity - importance hidden in complexity? Eur J Immunol 42(8):1969-1976.

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web URL: http://omim.org/ [Acessed August 2013].

Oostlander A, Meijer G, Ylstra B. 2004. Microarray-based comparative genomic hybridization and its applications in human genetics. Clin. Genet. 66:488-495.

Orth M, Schwenke C. 2011. Age-at-onset in Huntington disease. PLoS Currents 3.

Orth M. 2011. Observing Huntington's disease: the European Huntington's disease network's REGISTRY. J Neurol Neurosurg Psychiatry. 82:1409-1412.

Ou Z, Stankiewicz P, Xia Z, Breman AM, Dawson B, Wiszniewska J, Szafranski P, Cooper ML, Rao M, Shao L. 2011. Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. Genome Res. 21:33-46.

Pallos J, Bodai L, Lukacsovich T, Purcell JM, Steffan JS, Thompson LM, Marsh JL. 2008. Inhibition of specific HDACs and sirtuins suppresses pathogenesis in a Drosophila model of Huntington's disease. Hum Mol Genet. 17:3767-3775.

Pankratz N, Dumitriu A, Hetrick KN, Sun M, Latourelle JC, Wilk JB, Halter C, Doheny KF, Gusella JF, Nichols WC. 2011. Copy number variation in familial Parkinson disease. PloS One 6:e20988.

Pankratz N, Foroud T. 2007. Genetics of Parkinson disease. Genet. Med. 9:801-811.

Panov AV, Gutekunst C, Leavitt BR, Hayden MR, Burke JR, Strittmatter WJ, Greenamyre JT. 2002. Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. Nat. Neurosci. 5:731-736.

Parker JA, Connolly JB, Wellington C, Hayden M, Dausset J, Neri C. 2001. Expanded polyglutamines in caenorhabditis elegans cause axonal abnormalities and severe dysfunction of PLM mechanosensory neurons without cell death. Proc. Natl. Acad. Sci. U.S.A. 98:13318-13323.

Parker JA, Holbert S, Lambert E, Abderrahmane S, Néri C. 2004. Genetic and pharmacological suppression of polyglutamine-dependent neuronal dysfunction in Caenorhabditis elegans. J Mol Neurosci. 23(1-2):61-8.

Paulsen JS, Hayden M, Stout JC, Langbehn DR, Aylward E, Ross CA, Guttman M, Nance M, Kieburtz K, Oakes D. 2006. Preparing for preventive clinical trials: The predict-HD study. Arch. Neurol. 63:883.

Paulsen JS, Nopoulos PC, Aylward E, Ross CA, Johnson H, Magnotta VA, Juhl A, Pierson RK, Mills J, Langbehn D. 2010. Striatal and white matter predictors of estimated diagnosis for Huntington disease. Brain Res. Bull. 82:201-207.

Pereira C, Bessa C, Soares J, Leão M, Saraiva L. 2012. Contribution of yeast models to neurodegeneration research. J Biomed Biotechnol. 2012:941232.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R *et al.,* (x co-authors. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 39:1256-1260.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet. 20:207-211.

Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R. 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. 29:512-520.

Piotrowski A, Bruder CEG, Andersson R, de Ståhl TD, Menzel U, Sandgren J, Poplawski A, von Tell D, Crasto C, Bogdan A. 2008. Somatic mosaicism for copy number variation in differentiated human tissues. Hum. Mutat. 29:1118-1124.

Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Børresen-Dale AL, Brown PO. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc. Natl. Acad. Sci. U.S.A. 99:12963-12968.

Price C. 1993. Fluorescence in situ hybridization. Blood Rev. 7:127-134.

PrimeTime qPCR Assays available at :
http://eu.idtdna.com/scitools/Applications/RealTimePCR/ [Acessed August 2013].

Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. 2012. The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. Mov Disord. 27(9):1083-91.

Quarrell OW, Rigby AS, Barron L, Crow Y, Dalton A, Dennis N, Fryer AE, Heydon F, Kinning E, Lashwood A *et al*.,, (x co-authors. 2007. Reduced penetrance alleles for Huntington's disease: A multi-centre direct observational study. J. Med. Genet. 44:e68.

Ramos EM, Latourelle JC, Lee J, Gillis T, Mysore JS, Squitieri F, Di Pardo A, Di Donato S, Hayden MR, Morrison PJ. 2012. Population stratification may bias analysis of PGC-1α as a modifier of age at Huntington disease motor onset. Hum. Genet. 131(12):1833-40.

Ranen NG, Stine OC, Abbott MH, Sherr M, Codori A, Franz ML, Chao NI, Chung AS, Pleasant N, Callahan C. 1995. Anticipation and instability of IT-15 (CAG) n repeats in parent-offspring pairs with Huntington disease. Am. J. Hum. Genet. 57:593.

Raspe M, Gillis J, Krol H, Krom S, Bosch K, van Veen H, Reits E. 2009. Mimicking proteasomal release of polyglutamine peptides initiates aggregation and toxicity. J. Cell. Sci. 122:3262-3271.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:i333-i339.

Ravetch JV, Perussia B. 1989. Alternative membrane forms of fc gamma RIII (CD16) on human natural killer cells and neutrophils. cell type-specific expression of two genes that differ in single nucleotide substitutions. J. Exp. Med. 170:481-497.

Ravikumar B, Duden R, Rubinsztein DC. 2002. Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. Hum Mol Genet. 11:1107-1117.

Ravikumar B, Vacher C, Berger Z, Davies JE, Luo S, Oroz LG, Scaravilli F, Easton DF, Duden R, O'Kane CJ. 2004. Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. Nat Genet. 36:585-595.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W *et al*.,, (x co-authors. 2006. Global variation in copy number in the human genome. Nature 444:444-454.

Reekie, K. E. (2011). Technological and biological studies of human structural variation (Doctoral dissertation, University of Leicester).

Reilmann R. 2013. Pharmacological treatment of chorea in Huntington's disease-good clinical practice versus evidence-based guideline. Mov. Disord. 28:1030-1033.

Rein K, Zockler M, Mader MT, Grubel C, Heisenberg M. 2002. The Drosophila standard brain. Curr. Biol. 12:227-231.

Reiner A, Dragatsis I, Dietrich P. 2011. Genetics and neuropathology of Huntington's disease. Int. Rev. Neurobiol. 98:325-372.

Reymond A, Henrichsen CN, Harewood L, Merla G. 2007. Side effects of genome structural changes. Curr. Opin. Genet. Dev. 17:381-386.

Rigamonti D, Bauer JH, De-Fraja C, Conti L, Sipione S, Sciorati C, Clementi E, Hackam A, Hayden MR, Li Y. 2000. Wild-type huntingtin protects from apoptosis upstream of caspase-3. J. Neurosci.20:3705-3713.

Rigby WF, Wu YL, Zan M, Zhou B, Rosengren S, Carlson C, Hilton W, Yu CY. 2012. Increased frequency of complement C4B deficiency in rheumatoid arthritis. Arthritis & Rheumatism 64:1338-1344.

Robinson JI, Carr IM, Cooper DL, Rashid LH, Martin SG, Emery P, Isaacs JD, Barton A, Wilson AG, Barrett JH. 2012. Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis. Hum. Mutat. 33:741-749.

Rogaeva E, Kawarai T, George-Hyslop PS. 2006. Genetic complexity of Alzheimer's disease: successes and challenges. J. Alzheimer's Dis. 9:381-387.

Ross CA, Poirier MA. 2005. What is the role of protein aggregation in neurodegeneration? Nat Rev Mol Cell Biol. 6:891-898.

RoveletLecrux A, Campion D. 2012. Copy number variations involving the microtubule-associated protein tau in human diseases. Biochem. Soc. Trans. 40:672.

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M *et al.*„ (x co-authors. 2006. APP locus duplication causes autosomal dominant early-ons*et al*zheimer disease with cerebral amyloid angiopathy. Nat Genet. 38:24-26.

Rovelet-Lecrux A, Legallic S, Wallon D, Flaman JM, Martinaud O, Bombois S, Rollin-Sillaire A, Michon A, Le Ber I, Pariente J. 2011. A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. Eur J Hum Genet. 20:613-617.

Rørth P. 1996. A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. Proc. Natl. Acad. Sci. U.S.A. 93:12418-12422.

Rubinsztein DC, Leggo J, Chiano M, Dodge A, Norbury G, Rosser E, Craufurd D. 1997. Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. Proc. Natl. Acad. Sci. U.S.A. 94:3872-3876.

Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J, Chotai K, Connarty M, Craufurd D, Curtis A. 1996. Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am. J. Hum. Genet. 59:16.

Ryder E, Russell S. 2003. Transposable elements as tools for genomics and genetics in Drosophila. Brief Funct Genomics. 2:57-71.

Sancheti H, Akopian G, Yin F, Brinton RD, Walsh JP, Cadenas E. 2013. Age-dependent modulation of synaptic plasticity and insulin mimetic effect of lipoic acid on a mouse model of Alzheimer's disease. PloS One 8:e69830.

Saft C, Epplen JT, Wieczorek S, Landwehrmeyer GB, Roos RA, de Yebenes JG, Dose M, Tabrizi SJ, Craufurd D. 2011. NMDA receptor gene variations as modifiers in Huntington disease: A replication study. PLoS Currents 3.

Sarkar S, Perlstein EO, Imarisio S, Pineau S, Cordenier A, Maglathlin RL, Webster JA, Lewis TA, O'Kane CJ, Schreiber SL. 2007. Small molecules enhance autophagy and reduce toxicity in Huntington's disease models. Nat Chem Biol. 3:331-338.

Sathasivam K, Baxendale S, Mangiarini L, Bertaux F, Hetherington C, Kanazawa I, Lehrach H, Bates GP. 1997. Aberrant processing of the fugu HD (FrHD) mRNA in mouse cells and in transgenic mice. Hum Mol Genet. 6:2141-2149.

Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RL, Roos RA, Howland D. 2013. Aberrant splicing of HTT generates the pathogenic exon 1

protein in Huntington disease. Proc. Natl. Acad. Sci. U.S.A. 110:2366-2370.

Schiffer NW, Broadley SA, Hirschberger T, Tavan P, Kretzschmar HA, Giese A, Haass C, Hartl FU, Schmid B. 2007. Identification of anti-prion compounds as efficient inhibitors of polyglutamine protein aggregation in a zebrafish model. J. Biol. Chem. 282:9195-9203.

Schipper-Krom S, Juenemann K, Reits EA. 2012. The ubiquitin-proteasome system in Huntington's disease: are proteasomes impaired, initiators of disease, or coming to the rescue? Biochem Res Int. 2012:837015.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 30:e57-e57.

Schwartz DC, Cantor CR. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37:67.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525-528.

Seredenina T, Luthi-Carter R. 2012. What have we learned from gene expression profiles in Huntington's disease? Neurobiol. Dis. 45:83-98.

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet. 38:1038-1042.

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R. 2005. Segmental duplications and copy-number variation in the human genome. Am. J. Hum. Genet. 77:78.

Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R. 2008. A recurrent 15q13. 3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet. 40:322-328.

Shaw C, Li Y, Wiszniewska J, Chasse S, Zaidi S, Jin W, Dawson B, Wilhelmsen K, Lupski J, Belmont J. 2011. Olfactory copy number association with age at onset of Azheimer disease.

Neurology 76:1302-1309.

Shaw CJ, Lupski JR. 2005. Non-recurrent 17p11. 2 deletions are generated by homologous and non-homologous mechanisms. Hum. Genet. 116:1-7.

She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C. 2004. The structure and evolution of centromeric transition regions within the human genome. Nature 430:857-864.

Shin H, Kim MH, Lee SJ, Lee K, Kim M, Kim JS, Cho JW. 2013. Decreased metabolism in the cerebral cortex in early-stage Huntington's disease: A possible biomarker of disease progression? J Clin Neurol. 9(1):21-5.

Shulman JM, Chipendo P, Chibnik LB, Aubin C, Tran D, Keenan BT, Kramer PL, Schneider JA, Bennett DA, Feany MB. 2011. Functional screening of Alzheimer pathology genome-wide association signals in drosophila. Am. J. Hum. Genet. 88:232.

Singhrao S, Neal J, Morgan B, Gasque P. 1999. Increased complement biosynthesis by microglia and complement activation on neurons in Huntington's disease. Exp. Neurol. 159:362-376.

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R *et al.*„ (x co-authors. 2003. Alpha-synuclein locus triplication causes Parkinson's disease. Science 302:841.

Slack A, Thornton P, Magner DB, Rosenberg SM, Hastings P. 2006. On the mechanism of gene amplification induced under stress in escherichia coli. PLoS Genetics 2:e48.

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. 2003. Mutations in RAI1 associated with Smith–Magenis syndrome. Nat Genet. 33:466-468.

Smith R, Klein P, Koc-Schmitz Y, Waldvogel HJ, Faull RL, Brundin P, Plomann M, Li J. 2007. Loss of SNAP-25 and rabphilin 3a in sensory-motor cortex in Huntington’s disease. J. Neurochem. 103:115-123.

Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, Davies P, MacDonald ME, Gusella JF, Harper PS, Shaw DJ. 1993. Relationship between trinucleotide repeat expansion and

phenotypic variation in Huntington's disease. Nat Genet. 4:393-397.

Son JH, Shim JH, Kim K, Ha J, Han JY. 2012. Neuronal autophagy and neurodegenerative diseases. Exp. Mol. Med. 44:89-98.

Sosa MAG, De Gasperi R, Elder GA. 2012. Modeling human neurodegenerative diseases in transgenic systems. Hum. Genet. 131:535-563.

Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Spradling AC, Rubin GM. 1982. Transposition of cloned P elements into Drosophila germ line chromosomes. Science 218:341-347.

Squitieri F, Andrew S, Goldberg Y, Kremer B, Spence N, Zelsler J, Nichol K, Theilmann J, Greenberg J, Goto J. 1994. DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum Mol Genet. 3:2103-2114.

Squitieri F, Gellera C, Cannella M, Mariotti C, Cislaghi G, Rubinsztein DC, Almqvist EW, Turner D, Bachoud-Lévi A, Simpson SA. 2003. Homozygosity for CAG mutation in Huntington disease is associated with a more severe clinical course. Brain 126:946-955.

Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. Trends Genet. 18:74-82.

Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in disease. Annu. Rev. Med. 61:437-455.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE. 2008. Large recurrent microdeletions associated with schizophrenia. Nature 455:232-236.

Steffan JS, Agrawal N, Pallos J, Rockabrand E, Trotman LC, Slepko N, Illes K, Lukacsovich T, Zhu Y, Cattaneo E. 2004. SUMO modification of huntingtin and Huntington's disease pathology. Science 304:100-104.

Steffan JS, Bodaiē L, Pallosē J, Poelman M, McCampbellǵ A, Apostol BL, Kazantsev A, Schmidt E, Zhu YZ, Greenwald M. 2001. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila. Amino Acids 1:2,441.

Steinert JR, Campesan S, Richards P, Kyriacou CP, Forsythe ID, Giorgini F. 2012. Rab11 rescues synaptic dysfunction and behavioural deficits in a Drosophila model of Huntington's disease. Hum Mol Genet. 21:2912-2922.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848-853.

Strong TV, Tagle DA, Valdes JM, Elmer LW, Boehm K, Swaroop M, Kaatz KW, Collins FS, Albin RL. 1993. Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. Nat Genet. 5:259-265.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE. 2010. Diversity of human copy number variation and multicopy genes. Science 330:641-646.

Swaminathan S, Kim S, Shen L, Risacher SL, Foroud T, Pankratz N, Potkin SG, Huentelman MJ, Craig DW, Weiner MW. 2011. Genomic copy number analysis in Alzheimer's disease and mild cognitive impairment: An ADNI study. Int J Alzheimers Dis. 2011:729478.

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu X, Vincent JB, Skaug JL, Thompson AP, Senman L. 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. 39:319-328.

Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RA, Durr A, Craufurd D, Kennard C, Hicks SL, Fox NC, Scahill RI *et al*., (x co-authors. 2009. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: Cross-sectional analysis of baseline data. Lancet Neurol. 8:791-801.

Tabrizi SJ, Scahill RI, Durr A, Roos RA, Leavitt BR, Jones R, Landwehrmeyer GB, Fox NC, Johnson H, Hicks SL. 2011. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: The 12-month longitudinal analysis. The Lancet Neurology 10:31-42.

Taherzadeh-Fard E, Saft C, Andrich J, Wieczorek S, Arning L. 2009. PGC-1alpha as modifier of onset age in Huntington disease. Mol. Neurodegener 4:10.

Taherzadeh-Fard E, Saft C, Wieczorek S, Epplen JT, Arning L. 2010. Age at onset in Huntington's disease: Replication study on the associations of ADORA2A, HAP1 and OGG1. Neurogenetics. 11(4):435-9.

Tai YF, Pavese N, Gerhard A, Tabrizi SJ, Barker RA, Brooks DJ, Piccini P. 2007. Microglial activation in presymptomatic Huntington's disease gene carriers. Brain 130:1759-1766.

Tan L, Yu J, Tan L. 2012. The kynurenine pathway in neurodegenerative diseases: Mechanistic and therapeutic considerations. J Neurol Sci. 323(1-2):1-8.

Tang T, Chen X, Liu J, Bezprozvanny I. 2007. Dopaminergic signaling and striatal neurodegeneration in Huntington's disease. J. Neurosci.27:7899-7910.

Tartari M, Gissi C, Sardo VL, Zuccato C, Picardi E, Pesole G, Cattaneo E. 2008. Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. Mol. Biol. Evol. 25:330-338.

Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, Adam S, Greenberg C, Ives EJ, Clarke LA. 1994. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet. 6:409-414.

Tellez-Nagel I, Johnson AB, Terry RD. 1974. Studies on brain biopsies of patients with Huntington's chorea. J Neuropathol Exp Neurol. 33:308-332.

Teuling E, Bourgonje A, Veenje S, Thijssen K, de Boer J, van der Velde J, Swertz M, Nollen E. 2011. Modifiers of mutant huntingtin aggregation: Functional conservation of C. elegans-modifiers of polyglutamine aggregation. PLoS Currents 3.

Thevandavakkam MA, Schwarcz R, Muchowski PJ, Giorgini F. Targeting kynurenine 3-monooxygenase (KMO): Implications for therapy in Huntington's disease. CNS Neurol Disord Drug Targets. 9(6):791-800.

The Huntington's Disease Collaborative Research Group. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.

the Huntington's disease collaborative research group. Cell 72(6):971-9783.

Toffolatti L, Cardazzo B, Nobile C, Danieli GA, Gualandi F, Muntoni F, Abbs S, Zanetti P, Angelini C, Ferlini A. 2002. Investigating the mechanism of chromosomal deletion: Characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. Genomics 80:523-530.

Trask BJ, Massa H, Brand-Arpon V, Chan K, Friedman C, Nguyen OT, Eichler E, van den Engh G, Rouquier S, Shizuya H. 1998. Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. Hum Mol Genet. 7:2007-2020.

Trottier Y, Biancalana V, Mandel J. 1994. Instability of CAG repeats in Huntington's disease: Relation to parental transmission and age of onset. J. Med. Genet. 31:377-382.

Trushina E, Dyer RB, Badger JD, Ure D, Eide L, Tran DD, Vrieze BT, Legendre-Guillemin V, McPherson PS, Mandavilli BS. 2004. Mutant huntingtin impairs axonal trafficking in mammalian neurons in vivo and in vitro. Mol. Cell. Biol. 24:8195-8209.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D. 2005. Fine-scale structural variation of the human genome. Nat Genet. 37:727-732.

Twelvetrees AE, Yuen EY, Arancibia-Carcamo IL, MacAskill AF, Rostaing P, Lumb MJ, Humbert S, Triller A, Saudou F, Yan Z. 2010. Delivery of GABAARs to synapses is mediated by HAP1-KIF5 and disrupted by mutant huntingtin. Neuron 65:53-65.

Urban TJ, Weintrob AC, Fellay J, Colombo S, Shianna KV, Gumbs C, Rotger M, Pelak K, Dang KK, Detels R *et al*. (x co-authors. 2009. CCL3L1 and HIV/AIDS susceptibility. Nat. Med. 15(10):1110-1112.

UCSC Genome Browser database available at: http://genome.ucsc.edu [Acessed August 2013].

UniProt available at: http://www.uniprot.org/[Acessed August 2013].

Vacic V, McCarthy S, Malhotra D, Murray F, Chou H, Peoples A, Makarov V, Yoon S, Bhandari A, Corominas R. 2011. Duplications of the neuropeptide receptor gene VIPR2 confer

significant risk for schizophrenia. Nature 471:499-503.

Valenza M, Rigamonti D, Goffredo D, Zuccato C, Fenu S, Jamot L, Strand A, Tarditi A, Woodman B, Racchi M. 2005. Dysfunction of the cholesterol biosynthetic pathway in Huntington's disease. J. Neurosci.25:9932-9939.

van Ommen GJB. 2005. Frequency of new copy number variation in humans. Nat Genet. 37:333-334.

Veal CD, Xu H, Reekie K, Free R, Hardwick RJ, McVey D, Brookes A, Hollox EJ, Talbot CJ. 2013. Automated design of paralogue ratio test assays for the accurate and rapid typing of copy number variation. Bioinformatics. 29(16):1997-2003.

Veal CD, Reekie KE, Lorentzen JC, Gregersen PK, Padyukov L, Brookes AJ. 2013. A 129 kb deletion on chromosome 12 confers substantial protection against rheumatoid arthritis, implicating the gene SLC2A3. Hum. Mutat.

Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P. 2005. Position effects due to chromosome breakpoints that map approximately 900 kb upstream and approximately 1.3 mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet. 76(4):652-662.

Velier J, Kim M, Schwarz C, Kim TW, Sapp E, Chase K, Aronin N, DiFiglia M. 1998. Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways. Exp. Neurol. 152:34-40.

Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL. 2004. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. Mol. Cell 14:95-104.

Vienna Drosophila RNAi Center Stock available at: http://stockcenter.vdrc.at/control/main [Acessed August 2013].

Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL. 2003. End-sequence profiling: Sequence-based analysis of aberrant genomes. Proc. Natl. Acad. Sci. U.S.A. 100:7696-7701.

Vonsattel JPG, DiFiglia M. 1998. Huntington disease. Journal of Neuropathology & Experimental Neurology 57:369-384.

Vrijenhoek T, Buizer-Voskamp JE, Van Der Stelt I, Strengman E. 2008. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am. J. Hum. Genet. 83:504.

Wain LV, Armour JAL, Tobin MD. 2009. Genomic copy number variation, human health, and disease. The Lancet 374:340-350.

Walker FO. 2007. Huntington's disease. Semin. Neurol. 27:143-150.

Walker S, Janyakhantikul S, Armour JA. 2009. Multiplex paralogue ratio tests for accurate measurement of multiallelic CNVs. Genomics 93:98-103.

Wang K, Bucan M. 2008. Copy number variation detection via high-density SNP genotyping. Cold Spring Harbor Protocols 2008:pdb. top46.

Warby SC, Montpetit A, Hayden AR, Carroll JB, Butland SL, Visscher H, Collins JA, Semaka A, Hudson TJ, Hayden MR. 2009. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am. J. Hum. Genet.84:351-366.

Wellington CL, Ellerby LM, Hackam AS, Margolis RL, Trifiro MA, Singaraja R, McCutcheon K, Salvesen GS, Propp SS, Bromm M. 1998. Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. J. Biol. Chem. 273:9158-9167.

Wellington CL, Singaraja R, Ellerby L, Savill J, Roy S, Leavitt B, Cattaneo E, Hackam A, Sharp A, Thornberry N. 2000. Inhibiting caspase cleavage of huntingtin reduces toxicity and aggregate formation in neuronal and nonneuronal cells. J. Biol. Chem. 275:19831-19838.

Wexler NS, Lorimer J, Porter J, Gomez F, Moskowitz C, Shackell E, Marder K, Penchaszadeh G, Roberts SA, Gayan J *et al.*„ (x co-authors. 2004. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proc. Natl. Acad. Sci. U. S. A. 101:3498-3503.

Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubinstein S, Penney JB, Snodgrass SR, Shoulson I, Gomez F, Arroyo MAR. 1987. Homozygotes for Huntington's disease. Nature 326:194-197.

Weydt P, Pineda VV, Torrence AE, Libby RT, Satterfield TF, Lazarowski ER, Gilbert ML, Morton GJ, Bammler TK, Strand AD. 2006. Thermoregulatory and metabolic defects in Huntington's disease transgenic mice implicate PGC-1α in Huntington's disease neurodegeneration. Cell Metab. 4:349-362.

Weydt P, Soyal SM, Gellera C, Didonato S, Weidinger C, Oberkofler H, Landwehrmeyer GB, Patsch W. 2009. The gene coding for PGC-1alpha modifies age at onset in Huntington's disease. Mol. Neurodegener 4:3.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y, Makhijani V, Roth GT. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872-876.

Williams A, Sarkar S, Cuddon P, Ttofi EK, Saiki S, Siddiqi FH, Jahreiss L, Fleming A, Pask D, Goldsmith P. 2008. Novel targets for Huntington's disease in an mTOR-independent autophagy pathway. Nature Chemi Biol. 4:295-305.

Wong AM, Wang JW, Axel R. 2002. Spatial representation of the glomerular map in the Drosophila protocerebrum. Cell 109:229-241.

Xia J, Lee DH, Taylor J, Vandelft M, Truant R. 2003. Huntingtin contains a highly conserved nuclear export signal. Hum Mol Genet. 12:1393-1403.

Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP. 2008. Adaptive evolution of UGT2B17 copy-number variation. Am J Hum Genet. 83:337-346.

Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K *et al*.,, (x co-authors. 2007. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am. J. Hum. Genet. 80:1037-1054.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25:2865-2871.

Zeitlin S, Liu J, Chapman DL, Papaioannou VE, Efstratiadis A. 1995. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue.

Nat Genet. 11:155-163.

Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. 10:451-481.

Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet. 41:849-853.

Zhang X, Smith DL, Meriin AB, Engemann S, Russel DE, Roark M, Washington SL, Maxwell MM, Marsh JL, Thompson LM. 2005. A potent small molecule inhibits polyglutamine aggregation in Huntington's disease neurons and suppresses neurodegeneration in vivo. Proc. Natl. Acad. Sci. U. S. A. 102:892-897.

Zhang Y, Li M, Drozda M, Chen M, Ren S, Mejia Sanchez RO, Leavitt BR, Cattaneo E, Ferrante RJ, Hayden MR. 2003. Depletion of wild-type huntingtin in mouse models of neurologic diseases. J. Neurochem. 87:101-106.

Zheng Z, Diamond MI. 2012. Huntington disease and the huntingtin protein. Progress in Molecular Biology and Translational Science 107:189.

Zuccato C, Belyaev N, Conforti P, Ooi L, Tartari M, Papadimou E, MacDonald M, Fossale E, Zeitlin S, Buckley N. 2007. Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease. J. Neurosci.27:6972-6983.

Zuccato C, Cattaneo E. 2009. Brain-derived neurotrophic factor in neurodegenerative diseases. Nature Reviews Neurology 5:311-322.

Zuccato C, Ciammola A, Rigamonti D, Leavitt BR, Goffredo D, Conti L, MacDonald ME, Friedlander RM, Silani V, Hayden MR. 2001. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. Science 293:493-498.

Zordan MA, Cisotto P, Benna C, Agostino A, Rizzo G, Piccin A, Pegoraro M, Sandrelli F, Perini G, and Tognon G.. 2006. Post-transcriptional silencing and functional characterization of the Drosophila melanogaster homolog of human Surf1. Genetics. 172:229-241.

Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T. 2003. Huntingtin interacts with REST/NRSF to modulate the

transcription of NRSE-controlled neuronal genes. Nat Genet. 35:76-83.

Zuccato C, Valenza M, Cattaneo E. 2010. Molecular mechanisms and potential therapeutical targets in Huntington's disease. Physiol. Rev. 90:905-981.

Zwilling D, Huang SY, Sathyasaikumar KV, Notarangelo FM, Guidetti P, Wu HQ, Lee J, Truong J, Andrews-Zwilling Y, Hsieh EW. 2011. Kynurenine 3-monooxygenase inhibition in blood ameliorates neurodegeneration. Cell 145:863-874.