The dark side of the human genome: Translation of microsatellite mutations located in non-coding sequences in novel and toxic proteins in neurological diseases

Oculo-Pharyngo-Distal Myopathy (OPDM, OMIM #164310), Oculopharyngeal Myopathy with Leukoencephalopathy (OPML, OMIM #618637) and Neuronal Intranuclear Inclusion Disease (NIID, OMIM # 603472) are adult-onset genetic disorders characterized by variable clinical manifestations of the central nervous system (ataxia, tremor, dementia, etc.), associated with weakness and atrophy of the skeletal muscles of the face, the pharynx and the distal limbs. At the histopathological level, these disorders are characterized by the presence of intranuclear inclusions, which are ubiquitin- and p62-positive, but of unknown origin. Recently, the genetic causes of OPML, NIID and OPDM1 to 5 were identified as **identical expansions of 50 to 200 GGC repeats, however located within genetic regions annotated as non-coding** in at least 6 different genes: *LOC642361, LRP12*, *GIPC1*, *NOTCH2NLC*, *RILPL1* and *ABCD3.*

Here, we found that the GGC repeat expansions located in the lncRNA *LOC642361* and that causes OPML, within the 5'-untranslated regions of the *GIPC1* or *NOTCH2NLC* gene, responsible respectively of OPDM2 and NIID/OPDM3, or within the antisense transcript of *RILPL1* , which cause OPDM4, are all embedded in small, previously unrecognized ORFs. Consequently, these various GGC repeat expansions are translated in novel polyGlycine-containing proteins. Importantly, antibodies developed against these various proteins stain the typical p62-positive inclusions, confirming translation of these GGC repeat expansions into novel proteins in individuals with OPDM, OPML or NIID. Moreover, expression of these polyGlycine proteins is sufficient to form inclusions in muscle cells culture and in mouse models, leading to locomotor alterations associated with neurodegeneration and muscle fiber atrophy. Finally, we found a pharmaceutical compound that solubilize polyGlycine aggregates in cell cultures, thus bringing hope to develop therapeutical options for these neurological diseases.

Overall, these results highlight the complexity and richness of the human genome and the importance of yet unrecognized small ORFs in human pathology, notably for translation of microsatellite mutations in novel and pathogenic proteins.

Dr. Nicolas Charlet

Researches in my group (IGBMC, *https://www.igbmc.fr/en/recherche/teams/rna-diseases)* are mainly focused on "non-coding" genome and human diseases, first in cancer (PhD, University of Paris), then in genetic diseases (post-doctoral training, Baylor college of Medicine). I started my independent group on RNA diseases at the IGBMC (Strasbourg, France) in 2007. Since, we are studying how expanded repeats located in the "non-coding" parts of the genome lead to human genetic diseases. We are notably interested in Myotonic Dystrophy (DM) caused by expanded CTG or CCTG repeats, various neurological diseases (Fragile X-Associated Tremor/Ataxia Syndrome, Oculo-Pharyngo-Distal Myopathy, Neuronal Intranuclear Inclusion Disease, etc.) due to expanded CGG repeats and Amyotrophic Lateral Sclerosis and Frontotemporal Dementia (ALS-FTD) due to GGGGCC expansion in the first intron of the *C9ORF72* gene. These expanded repeats are pathogenic by diverse mechanisms, such as their non-canonical translation in toxic proteins and/ or sequestration of specific RNA-binding proteins. Our goals are (**i**) to elucidate the molecular causes of these diseases, (**ii**) to establish relevant cell and animal models of these diseases and (**iii**) to identify drugs able to correct pathogenicity in these models.