Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid lipofuscinosis

David E. Sleat [a,b,*], Erika Gedvilaite [c,d], Yeting Zhang [c,d], Peter Lobel [a,b], Jinchuan Xing [c,d,**]

[a] Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[b] Department of Biochemistry and Molecular Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[c] Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[d] Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:
Received 4 May 2016
Received in revised form 11 July 2016
Accepted 18 August 2016
Available online xxxx

Keywords:
Neuronal ceroid lipofuscinosis
Whole exome sequencing

ABSTRACT

The neuronal ceroid lipofuscinoses (NCLs) are a group of fatal, mostly recessive neurodegenerative lysosomal storage diseases. While clinically similar, they are genetically distinct and result from mutations in at least twelve different genes. Estimates of NCL incidence range from 0.6 to 14 per 100,000 live births but vary widely between populations and are influenced by whether patients are classified based upon clinical or genetic criteria. We investigated mutations in twelve NCL genes in ~61,000 individuals represented in the Exome Aggregation Consortium (ExAC) whole exome sequencing database. Variants were extracted from ExAC and pathogenic alleles were differentiated from neutral polymorphisms using annotated variant databases and missense mutation prediction tools. Carrier frequency was dependent on ethnicity, with the highest (1/75) observed for PPT1 in the Finnish. When data are adjusted for ethnic diversity within the USA, *PPT1*, *TPP1* and *CLN3* carrier frequencies were found to be the highest of the NCLs, each at ~1/500. Carrier frequencies calculated from ExAC correlated well with incidence estimated from numbers of living NCL patients in the US. In addition, the analysis identified numerous variants that are annotated as pathogenic in public repositories but have a predicted frequency that is not consistent with patient studies. These variants appear to be neutral polymorphisms that are reported as pathogenic without validation. Based upon literature reports, such alleles may be annotated in public databases as pathogenic and this propagates errors that can have clinical consequences.

1. Introduction

The neuronal ceroid lipofuscinoses (NCLs) are a group of mostly recessively-inherited neurodegenerative diseases that primarily affect children (Mole et al. 2011). The pathologic hallmark of these diseases is an accumulation of fluorescent material in the lysosomes of affected individuals and onset is typically marked by seizures and/or visual problems, which become progressively more severe, eventually accompanied by dementia and loss of locomotor function. Progression is relentless, and these diseases generally result in death. To date, defects in 14 different genes have been definitively associated with patients diagnosed with NCL disease (Table 1).

There are a number of estimates for the incidence of NCLs as a collective group in different European populations, and these range from ~0.6 (Italy) to ~14 (Newfoundland) per 100,000 live births (Table 1) (Claussen et al. 1992; Cardona and Rosati 1995; Mitchison et al. 1995; Crow et al. 1997; Elleder et al. 1997; Uvebrant and Hagberg 1997; Ostergaard and Hertz 1998; Rider and Rider 1999; Taschner et al. 1999; Teixeira et al. 2003; Augestad and Flanders 2006; Moore et al. 2008; Santorelli et al. 2013). For individual NCLs, studies have examined *PPT1* (previously denoted as *CLN1*), *TPP1* (previously denoted as *CLN2*) and *CLN3* and results are also population dependent (Table 1,). However, interpretation of epidemiological data for NCLs is complicated by the fact that many earlier studies were conducted before the identification of the respective disease genes, thus patients were defined by clinical criteria. While some NCLs can be accurately identified from the ultrastructure of the storage material, there is significant overlap in clinical presentation between forms with distinct genetic origins. In addition, to date, epidemiological studies have been confined to European or European-derived populations and little is known about the NCL prevalence and distribution in non-European populations.

Abbreviations: AFR, African; AMR, Latino; BDSRA, Batten Disease Support and Research Association; EAS, East Asian; ExAC, Exome Aggregation Consortium; FIN, Finnish; FTD, frontotemporal dementia; HGMD, Human Gene Mutation Database; NCL, neuronal ceroid lipofuscinosis; NFE, non-Finnish European; OTH, other; SAS, South Asian.

* Correspondence to: D.E. Sleat, Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.
** Correspondence to: J. Xing, Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.
E-mail addresses: sleat@cabm.rutgers.edu (D.E. Sleat), erika_gedvilaite@yahoo.com (E. Gedvilaite), yezhang@dls.rutgers.edu (Y. Zhang), lobel@cabm.rutgers.edu (P. Lobel), xing@biology.rutgers.edu (J. Xing).

http://dx.doi.org/10.1016/j.gene.2016.08.031
0378-1119/

Table 1
Description and reported incidence of NCL diseases.

Gene name and disease	Gene product	NCL type and alternate presentation	Incidence	Population	Defined	Reference
PPT1	Palmitoyl protein thioesterase 1	Infantile NCL	5 per 100,000	Finland	Clinical	Uvebrant and Hagberg, 1997
			0.16 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
			0.05 per 100,000	Czech Republic	Clinical	Elleder et al., 1997
TPP1	Tripeptidyl peptidase 1	Late-infantile, classical NCL and *spinocerebellar ataxia, autosomal recessive 7*	0.6–0.7 per million	Sweden, Norway, Finland	Clinical	Uvebrant and Hagberg, 1997
			0.46 per 100,000	West Germany	Clinical	Claussen et al., 1992
			0.36 per 100,000	Italy	Clinical	Cardona and Rosati, 1995
			0.15 per 100,000	Portugal	Genetic	Teixeira et al., 2003
			0.5 per 100,000	Netherlands	Clinical	Taschner et al., 1999
			0.62 per 100,000	Czech Republic	Clinical	Extrapolated from Elleder et al., 1997
			0.28 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
			9 per 100,000	Newfoundland	Genetic	Moore et al., 2008
CLN3	CLN3 protein	Juvenile NCL	0.5 per 100,000	Portugal	Genetic	Teixeira et al., 2003
			4.8 per 100,000	Finland	Clinical	Mitchison et al., 1995
			0.02 per 100,000	Czech Republic	Clinical	Extrapolated from Elleder et al., 1997
			1.6 per 100,000	Denmark	Clinical	Ostergaard and Hertz, 1998
			0.5 per 100,000	Newfoundland	Genetic	Moore et al., 2008
			0.15 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
			1.45 per 100,000	Netherlands	Clinical	Taschner et al., 1999
DNAJC5	DnaJ (Hsp40) homolog, subfamily C, member 5	Autosomal dominant adult NCL				
CLN5	CLN5 protein	Finnish variant late infantile, NCL	0.07 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
CLN6	CLN6 protein	Variant late infantile NCL	0.20 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
			0.62 per 100,000	Czech Republic	Clinical	Extrapolated from Elleder et al., 1997
MFSD8	Major facilitator superfamily domain containing 8	Variant late infantile NCL	2.6 per 100,000	Newfoundland	Genetic	Moore et al., 2008
			0.14 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
CLN8	CLN8 protein	Variant late infantile NCL and *Northern epilepsy*	0.07 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
CTSD	Cathepsin D	Congenital NCL	0.01 per 100,000	Italy		Extrapolated from Santorelli et al., 2013
GRN	Progranulin	Adult-onset NCL and aphasia, primary progressive, frontotemporal lobar degeneration with ubiquitin-positive inclusions				
ATP13A2	ATPase type 13A2	Juvenile-onset NCL				
CTSF	Cathepsin F	Autosomal recessive adult onset NCL				
KCTD7	Potassium channel tetramerization domain containing 7	Infantile-onset NCL				
SGSH	N-sulfoglucosamine sulfohydrolase	Adult-onset NCL and *MPS IIIA*				
Total NCL			13.6 per 100,000	Newfoundland	Genetic	Moore et al., 2008
Total NCL			13 per 100,000	Finland		Santavuori et al., 1974
			0.56 per 100,000	Italy	Clinical	Cardona and Rosati, 1995
			1.28 per 100,000	West Germany	Clinical	Claussen et al., 1992
			1.61 per 100,000	Western Scotland	Clinical	Crow et al., 1997
			1.95 per	Netherlands	Clinical	Taschner et al., 1999

Please cite this article as: Sleat, D.E., et al., Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid l..., Gene (2016), http://dx.doi.org/10.1016/j.gene.2016.08.031