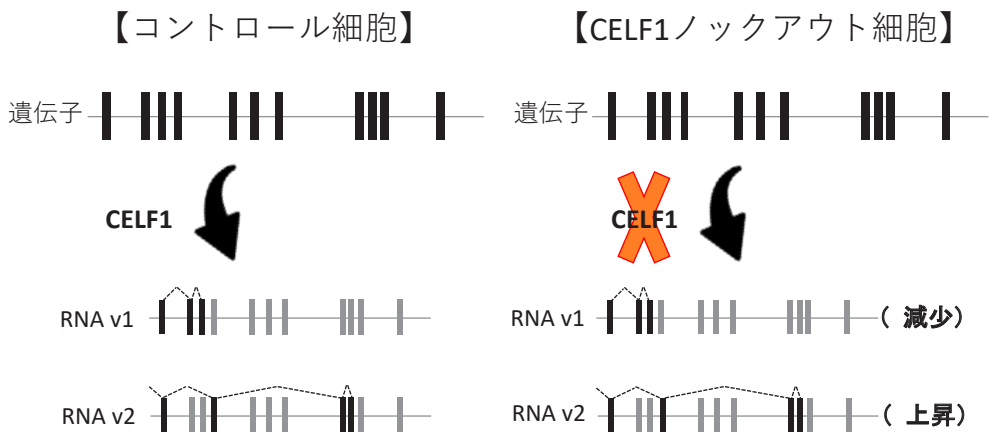
180　RNAプロファイリングによる認知症の発症機序解明	永田　健一

【目的】アルツハイマー病は高齢期発症の認知症の過半数を占める神経変性疾患であり、患者数の増加が深刻な問題となっている。疾患発症機序に関する理解は未だ限定的であり、根本原因を取り除くような治療薬は存在しない。ごく一部の遺伝性の患者では、原因遺伝子上の遺伝子変異が発症の引き金になる。また、孤発性患者においても、Genome-wide association study（GWAS）や次世代シークエンサーを用いて疾患発症の遺伝的素因について網羅的に探索することが可能になっており、これまでに孤発性アルツハイマー病の発症に関わる数十のリスク因子が同定されている。ただし、多くのリスク因子で機能的側面の知見は十分に集まっておらず、なぜアルツハイマー病発症のリスクとなるのかは不明瞭なままである。本研究ではアルツハイマー病の発症機序について新たな知見を得ることを目的とし、リスク因子である RNA 結合タンパク CELF1 についてゲノム編集技術や革新的なシークエンス技術を駆使して機能解析を行った。

【方法】まず、ゲノム編集技術 CRISPR/Cas9 を用いて *CELF1* ノックアウト細胞を樹立した。次に、未編集のコントロール細胞、ノックアウト細胞のそれぞれから RNA を抽出し、次世代シークエンサーを用いて RNA-seq を行った。網羅的探索の結果や過去の知見を踏まえて CELF1 の標的候補を絞り、さらに第三世代シークエンサーを用いて各転写産物の全長を定量的に評価した。

【結果】CELF1 の翻訳開始地点の周辺を狙って、CRISPR/Cas9 の発現プラスミドを細胞に導入した。ゲノム編集の効率を確認し、編集効率が高く、タンパク質の読み枠がずれる標的配列のみを以降の解析に使用した。プラスミド導入後、96 well プレートに 1 細胞ずつ入るように播き直し、細胞の増殖を待って DNA の編集結果をシークエンスで評価した。また、CELF1 抗体を使った western blotting で CELF タンパクが出ていないことを確認した。RNA-seq による解析で転写産物のスプライシングの異常を網羅的に探索し、HeLa 細胞および HEK 細胞のどちらの細胞株でも変化している遺伝子を CELF1 の標的候補とした。標的候補のリストの中で、*KLC1* 遺伝子についてはアルツハイマー病との関連が指摘されていたため、さらに第三世代シークエンサーを用いた全長の定量的解析を行った。*CELF1* ノックアウト細胞では KLC1 転写産物が質的・量的に変化していることが明らかとなった。

アルツハイマー病のリスク因子 CELF1 の機能解析



180. RNA プロファイリングによる認知症の発症機序解明

永田　健一

*大阪大学 大学院医学系研究科 認知症プレシジョン医療開発学寄附講座

Key words：アルツハイマー病，CELF1，ゲノム編集技術，次世代シークエンサー，第三世代シークエンサー

緒　言

超高齢社会の日本にとって認知症の克服は喫緊の課題である。認知症とは多数の疾患の総称であるが、中でも、神経変性疾患アルツハイマー病は高齢期発症型の半数以上を占めており、認知症の最大の原因となっている。アルツハイマー病の患者脳では約 40 アミノ酸のペプチドであるアミロイドβや微小管結合タンパク質タウの蓄積が認められる。これらの病理像は神経変性に先立って生じるために疾患発症に何らかの寄与があると考えられてきた。実際、家族性のアルツハイマー病患者の一部には、アミロイドβの代謝に関わる *APP*、*PSEN1*、*PSEN2* の 3 つが原因遺伝子として同定されている。タウをコードする遺伝子においても、その変異が神経変性疾患の発症の原因となることが報告されている。現在、世界規模で大型の臨床試験が多数進行中であり、そのほとんどはアミロイドβとタウタンパクを標的としている。また、原因遺伝子上の変異を導入することによりアルツハイマー病モデルマウスが作製され、研究に利用されている。

家族性患者から得られた様々な知見は極めて重要な位置を占めている。一方で、アルツハイマー病では、家族性に比べて孤発性の患者が圧倒的に多数であることも事実である。遺伝的素因が強力に働く家族性とは異なり、孤発性の例では、遺伝的素因に加えて生活習慣や加齢などの様々な要因が複雑にからみあって疾患発症に至る。解析技術の発展に伴って、孤発性の患者についても遺伝的素因がどの程度発症に寄与するのか知見の蓄積が加速することとなった。数万人規模の参加者を対象とした Genome-wide association study（GWAS）等により、これまでに多数の遺伝的多型が同定されている［1, 2］。多くの場合において、同定された多型の疾患発症リスクは高くないものの、その保有頻度は原因遺伝子変異に比べて桁違いに高い。従って、ヒト集団がどのようにアルツハイマー病という疾患を発症するかを理解する上で、多型と紐づけられた疾患リスク因子の機能を解明することは重要である。

CUGBP Elav-like family member 1（以下 *CELF1*）は 2013 年の GWAS の報告により、アルツハイマー病との関連が指摘されたリスク因子の 1 つである。*CELF1* は RNA 結合タンパクをコードする遺伝子であり、これまでにスプライシング制御や転写産物の安定化、翻訳制御への関与が報告されている。CELF1 は胎生期から全身性に発現することから、骨格筋、心筋その他各種臓器の発達へ関与していると考えられてきた。実際、ノックアウトマウスは体が小さく、生後の致死率が高いこと、精子形成に異常が出ることが分かっている［3］。また、ヒトのゲノム情報を集約したリソース gnomAD で検索してみると、遺伝子機能喪失を引き起こすような CELF1 変異の頻度は予想されるよりもはるかに低くなっている。ヒト集団内での選択は、CELF1 がヒトの生存にも重要な役割を担っている可能性を示唆している。近年、CELF の機能を明らかにするために、CELF1 に結合する RNA を回収し、次世代シークエンサーで網羅的に塩基配列を解読する試みが実行されてきた［4，5］。文献ごとに数千の転写産物が得られているものの、研究結果が文献ごとに異なっており、真の標的を捉えているのかどうかは不明瞭である。また、短い塩基配列を解析対象とする次世代シークエンサーのみで解析していることにも欠点がある。遺伝子の全長にわたって生じるスプライシングの変化を精度高く同定するには、別の技術を使った解析が必要である。

本研究ではリスク因子 CELF1 の機能解析をすることで、アルツハイマー病の発症機序についての新たな知見を得ることを目的とした。先行研究間の結果が異なる、という課題を克服するため、2 種類の培養細胞（HeLa と HEK293T）で同一の解析を実行し、同じ傾向が得られたものを標的候補とした。さらに、標的候補の全長を捉えるために第三世代シークエンサーを駆使した解析を併せて行った。

*現在の所属：名古屋大学 大学院医学系研究科 機能組織学

方　法

1.　ゲノム編集実験

CELF1 の開始コドン周辺を標的として CRISPR/Cas9 の標的配列を設計した。まず CRISPOR による標的配列の検索を行い、算出されるスコアが高いものを標的配列候補として残した（CRISPOR http://crispor.tefor.net/）。次に、別の検索ツール inDelphi でフレームシフト効率が高いと判定された 5 つを標的配列として選択した（inDelphi https://indelphi.giffordlab.mit.edu/）。対応したオリゴを合成し、プラスミド（Addgene #62988）への組み込みを行った。構築した発現プラスミドは大腸菌に入れて増やした後、市販のキットを使って回収した。シークエンス解析を使って発現プラスミドが適切に構築されているかを確認し、培養細胞への導入用に高純度のものを回収した。培養細胞へのプラスミド導入は Lipofectamine LTX を用いてリバース法で行った。24 well プレートを使用し、細胞密度は 5×10^4/well とした。24 時間後に細胞から DNA を抽出し、Heteroduplex Mobility Assay で編集の有無を検証した。

編集が効率よく導入される標的配列を使って *CELF1* ノックアウト細胞を作製した。上記と同様の手順でプラスミドを導入し、一部を回収、編集効率をチェックした。その後、96 well プレートに well あたり 1 細胞になるよう播種した。細胞の増殖を待った後、*CELF1* 遺伝子領域のシークエンス解析を行った。読み枠のずれによって *CELF1* 遺伝子の機能が欠失すると予想された細胞を解析対象の候補として残した。また、細胞からタンパク質を回収し、western blotting を行った。抗体は以下のものを使用した。CELF1 抗体（Millipore、05-621、Anti-CUG-BP1 mouse monoclonal、clone 3B1）、β-Actin 抗体（SIGMA、A5441-100UL Monoclonal Anti-β-Actin antibody）、2 次抗体（PROMEGA、W402、Anti-Mouse IgG (H＋L)、HRP Conjugate）。希釈倍率は CELF1 が 5,000 倍、β-Actin が 20,000 倍、2 次抗体が　10,000 倍で使用した。

HEK293T 細胞、HeLa 細胞の 2 種類について、同様の手順でノックアウト細胞株を複数樹立した。

2. RNA-seq 解析

ノックアウト細胞、および未編集のコントロール細胞を同じ細胞数回収し、RNA を抽出した。CRISPR/Cas9 の発現プラスミド導入は行わず、シングルセルクローニングのみを実行した細胞を未編集の比較対象とした。HEK293T 細胞のサンプル数はノックアウト細胞、コントロール細胞のそれぞれで 11 サンプル、HeLa 細胞ではそれぞれ 7 サンプルであった。定法に従ってライブラリー調整を行った後、illumina の装置を使って 100 bp ペアエンド、サンプルあたり 3,000 万リードで RNA-seq を行った。取得されたシークエンスは、リファレンス配列にマッピングした後、Cuffdiff などを用いて発現量の定量的な解析を行った。また、DEXSeq［6］を用いて exon ごとの相対的なリード数を定量することでスプライシングの変化を評価した。発現量とスプライシングの変化についてコントロールとノックアウト間で比較解析した。

3. 転写産物全長の解析

ノックアウト細胞、および未編集のコントロール細胞から抽出した RNA を逆転写し、PCR により *KLC1* 遺伝子領域を増幅した。*KLC1* の転写産物は 3 末端の非翻訳領域が多様性に富むことが分かっていたため、網羅的に転写産物を取得できるよう、3 末端側のプライマーは逆転写反応時に付加したアダプター上に設計した。サンプル数はノックアウト細胞、コントロール細胞のそれぞれで 6 サンプルずつの計 12 サンプルとした。定法に従ってライブラリー調整を行った後、Oxford Nanopore Technologies 社の第三世代シークエンサーMinION を使って 1 分子ずつ全長の配列を取得した。シークエンス解析は UNAGI［7］を使って行った。また、HEK293T 細胞、HeLa 細胞で同様の実験を行った。

結　果

1. *CELF1* ノックアウト細胞の樹立

先行文献［4, 5］で実施されている CELF1 結合 RNA の網羅的探索は、有用なスクリーニング手法である。ただし、結合した結果、RNA がどのような変化を引き起こすか、という機能的側面は分からない。先行文献［5］では siRNA

を使ったノックダウン実験を併せて行ってはいるものの、サンプル数が 2 と少なかった。また、siRNA は遺伝子機能を調べるための簡便なツールではあるものの、ノックダウンは一過性であり、導入量によっては細胞毒性が生じることも知られている。そこで本研究では、CRISPR/Cas9 でノックアウト細胞を多数作製し、未編集のコントロール細胞と細胞中の RNA の量的・質的変動を比較する機能的スクリーニングを実行することにした。

コンピュータ上で設計した標的配列を細胞に導入して編集効率をみたところ、5 つ中 4 つの標的配列では、電気泳動後にバンドのシフトが検出され、ゲノム編集が生じていることが示唆された。編集後の細胞から DNA を抽出し、標的配列周辺を増幅した後、サンガー法にてシークエンスを確認したところ、期待したように読み枠がずれている細胞が多数とれていた。さらに、次世代シークエンサーを使って定量的に評価し、4 つ中、2 つの標的配列（CELF1 sgRNA1、CELF1 sgRNA2）で 1 塩基の挿入欠失が高い頻度で生じることが分かった。CELF1 sgRNA1 あるいは CELF1 sgRNA2 を導入後に 1 細胞ずつに播き直し、増殖を待って DNA を抽出した。サンガー法によるシークエンス解析を行い、読み枠がずれているものをノックアウト細胞として残した。ピークが多数存在するために、シークエンスの読み取りが困難な場合は、TIDE（https://tide.nki.nl/）を使うことでコンピュータ的に処理し、混在する genotype を識別した。DNA レベルで 90%以上ノックアウトされている細胞については、さらにタンパク質を抽出し、CELF1 の western blotting を行った。未編集の細胞では、予想分子量付近にバンドが検出されたが、ノックアウト細胞では消失していた。一方で、インターナルコントロールである β-Actin ではノックアウト細胞でも未編集細胞と同様のバンドが検出された。DNA レベルだけでなくタンパクレベルでも機能の欠失が予想される細胞のみをノックアウト細胞と定義した（HEK293T で 11 細胞、HeLa 細胞で 7 細胞）。

2. 次世代シークエンサーによる解析

HEK293T と HeLa の 2 種類について、次世代シークエンサーを使ってノックアウト細胞とコントロール細胞の RNA-seq を行った。HEK293T 細胞では、ノックアウト細胞とコントロール細胞間で 2,017 遺伝子の発現量に有意差がついた。HeLa 細胞では 2,247 遺伝子であった。多重比較となるため、有意水準は補正後の値で 0.05 とした。2 種類での重複をカウントすると 313 遺伝子であった。これらの遺伝子について発現変動量を散布図にプロットしたところ、HEK293T と HeLa で変動量に相関が認められた（r=0.46）。重複した遺伝子群についてエンリッチメント解析を行ってみると、「regulation of neuron differentiation」などいくつかがヒットした。

スプライシングについては、HEK293T で 1,446、HeLa で 3,772 箇所に変化が認められた。重複は 322 箇所であった。有意水準は DEXSeq のデフォルト 0.1 のまま使用した。変動が認められた遺伝子の中には、*CELF1* 自身も含まれていたが、ゲノム編集による影響が強く反映されていると考え除外した。変動量をプロットしてみると、HEK293T と HeLa で高い相関が認められた（r=0.78）。重複して変動したのは計 151 遺伝子であった。1 箇所でしか変動していない遺伝子もあれば、複数の箇所で変動している遺伝子も存在していた。

3. 第三世代シークエンサーによる解析

次世代シークエンサーは 100 bp ほどの短い配列のみが解析対象であるため、複数の箇所でスプライシングの変動がみられた遺伝子については、全長の配列を正確に検出するのは困難である。そこで、Oxford Nanopore Technologies 社の第三世代シークエンサーMinION を用いて、転写産物全長の同定を行うことにした。*KLC1* の転写産物は、14 箇所で変動が確認されていただけでなく、過去の知見［8, 9］からアルツハイマー病への関与が疑われたために本研究で解析対象とした。RNA-seq の結果の妥当性は、複数の *KLC1* 遺伝子領域を標的とした real time PCR によっても支持された。細胞からの RNA 抽出後、*KLC1* 全長を増幅し、ライブラリー化の後にシークエンスした。次世代シークエンサーでは区別困難であった転写産物全長を取得することができた（図 1）。

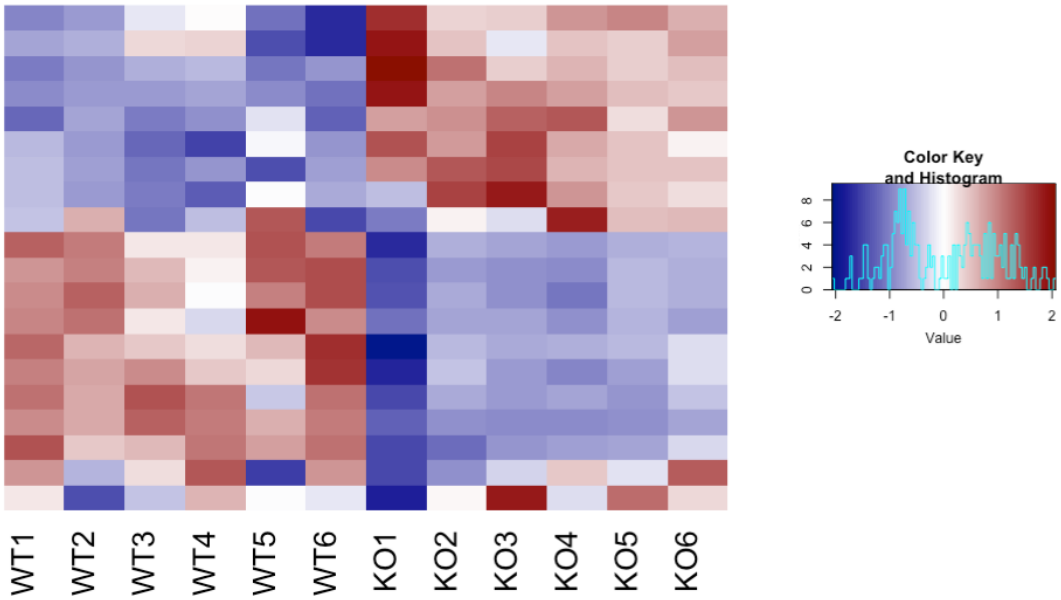


図 1. *CELF1* ノックアウト細胞における *KLC1* 転写産物の変化

縦軸に *KLC1* 転写産物、横軸に細胞サンプルをとり、ヒートマップで結果を表現した。発現量は相対的なものであり、上昇を赤、減少を青で表している。WT では上半分の転写産物が KO に比べて相対的に低かった。逆に、下半分の転写産物はおおむね KO の方が低い発現量であった（未発表データ）。

考　察

本研究ではアルツハイマー病のリスク因子である CELF1 についてゲノム編集技術とシークエンス解析を駆使して機能解析を行った。これまでにも培養細胞系での網羅的な標的 RNA の探索は行われていたが、先行研究間で結果が一致していなかった。また、先行研究では次世代シークエンサーで探索が行われるなど、基本的に‘short-read’の解析手法に留まっており、スプライシングという転写因子の全長にまたがって生じる現象について正しく評価できているかが不明であった。本研究では 2 種類の細胞を使って同じ手順で実験を行い、2 種類に共通してみられた変化をより信頼性が高い結果として扱った。また、偶然による結果の混入を避けるために、RNA-seq の解析サンプル数を十分に確保した。また、RNA-seq のスプライシング解析で顕著に差が見られた *KLC1* 遺伝子についてはさらに第三世代シークエンサーを用いて全長の変化を解析した。*CELF1* ノックアウト細胞における *KLC1* の変化は、RNA-seq のみでなく、定量的 qPCR、第三世代シークエンサーの全てで一致した結果が得られた。また、細胞を変えても類似した結果が認められた。

本研究では 1 年間という期間的な制約のため、当初の目的であったアルツハイマー病患者脳における解析については十分な時間を確保することが難しかった。第三世代シークエンサーによる *KLC1* 転写産物の全長解析は、死後脳サンプルにおいてもパイロット実験を行っており、現在、結果を精査しているところである。ただし、培養系とは異なって *CELF1* の完全ノックアウトではないこと、多様な細胞集団が混在していること、などから解析はスムーズに進まない可能性も十分にある。CELF1、KLC1 ともに脳の様々な細胞で発現しているため、CELF1 が KLC1 のスプライシング制御に寄与しているかを証明するには、組織中に含まれる細胞構成比率の影響を除く必要がある。第三世代シークエンサーのみの単純な解析では影響を除くことは困難であり、情報解析的なアプローチ、シングルセル解析によるアプローチなど課題克服のためのさらなる取り組みが必要である。

CELF1 が標的とする RNA を同定するために HEK293T、HeLa を用いた。細胞によって標的が異なることは十分に想定されるので、本研究でヒットした遺伝子群が脳においても同様に CELF1 の標的となるかどうか検証が必要

である。同じ培養細胞でも、例えば神経系由来の細胞株を使った方が結果の解釈が容易であった。研究遂行の中で SH-SY5Y などの神経系由来の細胞株でも遺伝子導入を行ったが、導入効率の低さやシングルセルクローニング時の形態変化などの問題が生じたため、効率を重視して HEK293T、HeLa の 2 種類に絞った。

KLC1 の多様なスプライシングがなぜ必要なのか、は本研究から明らかでない。軸索輸送に関わる KLC1 は特に神経細胞での役割が想定されるが、一部のアミノ酸配列異なることでどのような機能的差異が生じるのかさらなる解析が必要である。過去の報告から想定されたように、CELF1 は多数の遺伝子を標的としており、ノックアウト細胞では、発現量やスプライシングで様々な遺伝子に変化が認められた。KLC1 と同様、これらの変化についても機能的意義を知るためにはさらなる解析が必要である。

いわゆる遺伝子（protein coding gene）はタンパク質として機能する。どのようなタンパク質が産生されるかは転写産物全長を解読するのが 1 つの手段である。一方、次世代と比較すると第三世代シークエンサーのスループットは低くいとされる。本研究で採用したような、次世代シークエンサーによる 1 次スクリーニング、第三世代による 2 次スクリーニングは、両方のシークエンサーの長所が生かせるため、スプライシング因子の機能解析などで有効と考えられる。

共同研究者・謝辞

本研究の RNA-seq 解析は大阪大学微生物病研究所の元岡大祐助教の協力を得て行った。また、第三世代シークエンサーの解析は、元岡助教、同所属の大学院生 Mohamad Al kadi の支援を受けて実施した。

文　献

1) Jean-Charles Lambert et al., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics 2013 Dec;45(12):1452-8. doi: 10.1038/ng.2802. Epub 2013 Oct 27

2) Brian W. Kunkle et al., Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aß, tau, immunity and lipid processing. Nature Genetics 2019 Mar;51(3):414-430. doi: 10.1038/s41588-019-0358-2. Epub 2019 Feb 28.

3) Chantal Kress et al., Inactivation of CUG-BP1/CELF1 Causes Growth, Viability, and Spermatogenesis Defects in Mice. Mol Cell Biol. 2007 Feb;27(3):1146-57. doi: 10.1128/MCB.01009-06. Epub 2006 Nov 27.

4) Olivier Le Tonqueze et al., Identification of CELF1 RNA targets by CLIP-seq in human HeLa cells. Genomics Data. 2016 Apr 19;8:97-103. doi: 10.1016/j.gdata.2016.04.009. eCollection 2016 Jun.

5) Heng Xia et al., CELF1 preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2017 Sep;1860(9):911-921. doi: 10.1016/j.bbagrm.2017.07.004. Epub 2017 Jul 19.

6) Simon Anders et al., Detecting differential usage of exons from RNA-seq data. Genome Research. 2012 Oct;22(10):2008-17. doi: 10.1101/gr.133744.111. Epub 2012 Jun 21.

7) Mohamad Al kadi et al., UNAGI: an automated pipeline for nanopore full-length cDNA sequencing uncovers novel transcripts and isoforms in yeast. Functional & Integrative Genomics. 2020 Jan 18. doi: 10.1007/s10142-020-00732-1. [Epub ahead of print]

8) Gorazd B. Stokin et al., Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. Science. 2005 Feb 25;307(5713):1282-8. doi: 10.1126/science.1105681.

9) Takashi Morihara et al., Transcriptome analysis of distinct mouse strains reveals kinesin light chain-1 splicing as an amyloid-ß accumulation modifier. Proc Natl Acad Sci U S A. 2014 Feb 18;111(7):2638-43. doi: 10.1073/pnas.1307345111. Epub 2014 Feb 4.