

平成 29 年度老人保健事業推進費等補助金
老人保健健康増進等事業

認知症の診断および介護に対する AI および
IoT の活用に関する探索的調査研究

報告書

平成 30 (2018) 年 3 月

国立研究開発法人 国立長寿医療研究センター

平成 29 年度老人保健健康増進等事業 報告書

「認知症の診断および介護に対する AI および IoT の活用に関する探索的調査研究」報告書

目 次

[健康長寿支援ロボットセンター 近藤和泉 加藤健治]

AI を用いた認知症患者の排尿行動解析から生活支援機器の IoT 化を図る研究..... 1

[メディカルゲノムセンター 尾崎浩一 新飯田俊平]

1. 全ゲノム関連解析の施行、解析及び全ゲノム、全 RNA 配列解析の準備と検討..... 18

2. miRNA 発現情報と臨床情報の整備及び深層学習による認知症タイプの判別法の検討.....25

3. 参考資料 日本人類遺伝学会第 62 回大会 発表資料（平成 29 年 11 月 17 日） 28

平成 29 年度老人保健健康増進等事業 報告書

事業名: 認知症の診断および介護に対する AI および IoT の活用に関する探索的調査研究

国立研究開発法人 国立長寿医療研究センター

健康長寿支援ロボットセンター 近藤和泉 加藤健治

AI を用いた認知症患者の排尿行動解析から生活支援機器の IoT 化を図る研究

1. 研究背景および研究目的

認知症患者が夜間に尿意を感じた場合、その排尿行動は例え自身のバランス障害が存在しても、ためらいなく遂行される。日中であれば家族の目があり、また明るい場所で移動が行われることもあって、比較的問題は少ない。しかし夜間では状況が一変する。睡眠からの覚醒が十分では無く、暗い部屋や廊下をトイレまで移動することで、転倒リスクは日中と比べて驚異的に跳ね上がってしまう。このため、認知症患者の排尿行動時の転倒リスクを管理するための生活支援機器の IoT 化は喫緊の課題である。この問題を解決するため、我々は以下の序列での研究が必要と考えた。課題 1) 認知症患者の比率が高い当センターの入院患者における転倒インシデントの分析を通じて、尿意に発する行動が転倒リスクをどの程度、高くするかを検討、課題 2) ベッド上からトイレに向かう排尿行動が起こる前に、尿意を検知するシステムの開発のための基礎的なトライアルと計測データの AI による分析、課題 3) 早期の検知のために、尿意を感じた時に起こる自律神経系の変化の特徴抽出、課題 4) その特徴を抽出して尿意検知が可能になった場合の IoT を使った生活支援システムの構築。現状、上記課題の 2) には着手しており、マイクロウェーブを利用した呼吸数、心拍数の計測は可能となっている。しかし、主に通信環境の問題から、病院内ではマイクロウェーブによる計測データが安定して得られる環境が構築できないため、同データの AI による分析までには至っていない。このため、今回の研究では課題 1) に AI を利用した検討を行った。同時に課題 4) にも着手しており、排尿行動に移る前のストップングを行うための傾聴ロボット、トイレまでの移動時に安全を確保する杖・歩行器ロボットの開発にも取り組んでいる。また課題 3) に関して

は、実験系の構築を行い、来年度には着手し、その解析にも AI を使う予定としている。

さて認知症に限らず我が国において、超高齢化にともなう高齢者の転倒・転落（以後、転倒とする）への対応は、もっとも重大かつ急務の課題の一つである。さらに、高齢化の進展とともに、認知症患者も 2025 年には約 700 万人まで急増する見込みで、これまで研究領域としてスポットを浴びてこなかった認知症における転倒リスクを明らかにし、それに対応するための対策・環境整備が急務であると考ええる。

申請者らが実施した予備研究のデータより、高齢者の転倒原因となる行動要因として最も多いものが「尿意」であった。本研究では、この予備調査を発展させ、認知症患者における内的要因（行動意図、認知機能、転倒歴、年齢、性別等）や、外的要因（転倒場所、環境等）とを結び付けた、さらに詳細な転倒リスクを、統合的かつ複層的に調査することを目的とする。具体的には、日本 IBM が開発した人工知能 (Artificial intelligent: AI) “Watson Explorer” を活用して、当センターでこれまで 3000 件以上を集積してきた認知症入院患者における転倒転落事故報告書の大規模テキストマイニング解析を試みる。それによって、認知症患者の行動意図等の内的要因、更には、転倒場所等の外的要因間の転倒関連因子間強度を明らかにし、これまでアプローチできなかった認知症患者における転倒リスクを、内的・外的要因間で複層的かつ統合的に検討することを目的とする。

2. 研究方法

本研究では、日本 IBM との共同研究により、日本 IBM が開発した AI “Watson Explorer” を用いて、転倒転落報告書のテキストマイニングを行う。それによって、当センターでこれまで集積した約 3000 件を超える大規模転倒転落報告書のテキストマイニング解析を試み、患者の内的要因の一つである行動意図因子(例:トイレ、歯磨き、廊下の移動等)に絞った因子分析をする。それらを達成するために、初めに、患者の行動意図因子の中でも“尿意”に関連した因子を同定するための新規アルゴリズムを検討する。

2. 1. テキスト解析における課題 – 因子フラグ付け

初めに、大規模なテキスト解析を行う際の課題と、その課題を克服するための因子フラグ付け解析について述べる。本研究で用いるような大規模な自由記述テキストデータを解析対象にする際、単純な単語検索では、前後の単語や同義語を加味した因子抽出ができない。因子抽出/因子構造化とは、ある概念の定義を満たす表現がテキストに含まれるか否かをフラグ付けすることを指す。今回は、除外語や同義語の組み合わせを修正(チューニング)することにより、因子の定義に合致する文章を選択的に抽出することを試みた。下記に、除外語を含む表現からフラグを除外する例と、同義語を抽出する例を紹介する。

例1: 除外語を含む表現からフラグを除外

- ✓ 「トイレに行くために歩行して転倒。」
→ 排尿意図あり
- ✓ 「他患者のトイレ介助のために訪室し」
- ✓ トイレではなく食堂に移動中だったとのこと。
→ 排尿意図なし

* 「トイレ」という単語はあるが、「他患者の」、「～ではなく」とあるので、排尿意図による歩行では

ないと判断。

例2: 同義語の抽出

✓ 「便座の近くでズボンを履こうとして転倒」

→ 排尿意図 あり

✓ 「ポータブル使用のためベッドから立ち上がろうとして転倒」

→ 排尿意図 あり

* 「便座」と「ポータブル」はいずれもトイレの同義語なので、どちらのケースも排尿意図ありと判断。

2. 2. フラグ付けの精度向上方法詳細 – 正確性と網羅性のトレードオフについて

除外語と採用語の組み合わせをチューニングすることで因子抽出の精度を向上させることができるものの、正確な抽出と抽出漏れの防止はトレードオフ関係にある。正確性と網羅性のトレードオフをより具体的に記述すると、除外語を増やし、条件を厳しくすることで定義に合っている文章のみを抽出すると抽出漏れが生じる一方で、採用語を増やして幅広く抽出すると誤った文章を多く抽出してしまうことを指す。そこで、これらのトレードオフの両方の視点を組み合わせた指標である F-measure を用いることで、チューニング結果の精度を総合的に算出した(図1)。

本研究では、排尿意図に関連する単語が含まれるレコードの件数をまず算出し、その後同義語の追加・除外語の追加を行うことで、F-measure を向上させるという方針で行った(図2)。

2. 3. フラグ付けの精度向上方法詳細 – 因子抽出の精度検証方法について

最後に、排尿意図に起因する事故に該当する文章の因子定義を行い、定義に合致する文章のみにフラグ付けがされるよう除外語と採用語を組み合わせでチューニングを行った。具体的な因子定義とそのルールについては、図3に記載した。

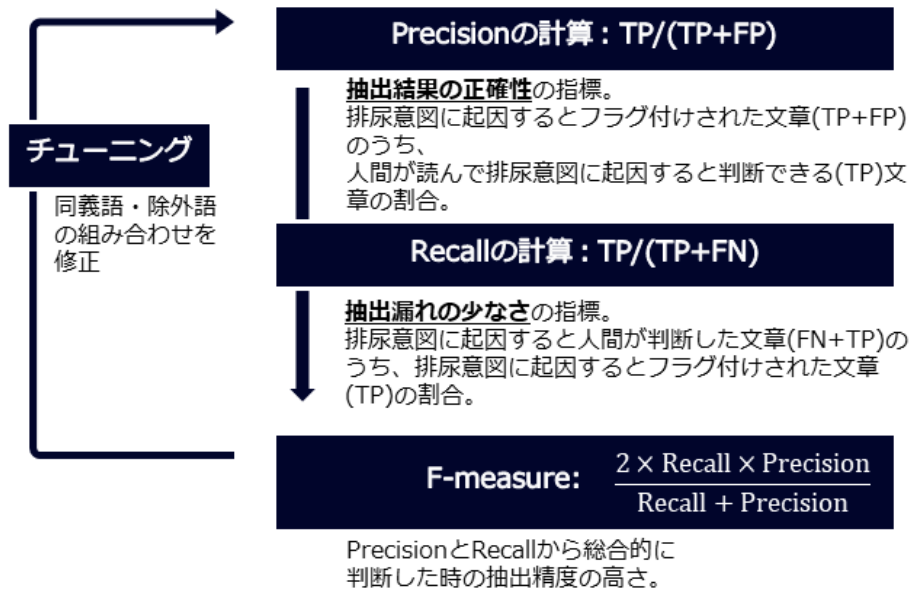


図1. 正確性と網羅性の両方の視点を組み合わせた指標である F-measure の算出方法

	概要	フラグ付けされる文章の例	レコード数	F-measure
1	単語検索 「排尿意図」に関連する単語による検索 同義語：トイレ、ポータブル、用足し、尿器など10件(プロジェクト開始時に定義)	- 「トイレに行こうとして転んだ」と話される。 - 他患者のトイレ介助のため訪問すると、マットの上に倒れているのを発見。	1577/3439	F-measure : 0.87 Precision : 0.80 Recall : 0.96
2	同義語追加 本文に含まれる名詞を全て抽出し、排尿意図に関連がありそうなものを追加 追加同義語：花、便所、尿失禁など9件(詳細は後述)	上記2例に加え - 「お花畑に花を・・・。」と言う。(認知症) - 何をしようとして、吸収マット上に降りたのか確認したところ「便所」と返答あり	1730/3439	チューニング による 改善
3	除外語追加 「排尿意図により転倒事故が起こった」とは言えない文章からフラグを除外 除外条件: 否定語を含む/主語が異なる文章	- 他患者のトイレ介助のため訪問すると、マットの上に倒れているのを発見。 →フラグ削除	1541/3439	
4	チューニング後 除外語をさらに追加し、定義に合わない文章からフラグ削除 除外条件: 「普段」に関する表現	- 普段はトイレからナースコールがある。 →フラグ削除	1425/3439	

図2. 本研究における行動意図の抽出方針に関する概要図

因子定義

	方針	例文
フラグ付けを行う	トイレに向かう際に発生した事故。	トイレに行く際ふらついて転倒。
	トイレから戻る際の事故(寄道は含まず)。	トイレから戻ろうとして転倒。
	排尿行為が完了するまでに発生した事故	排尿後尿器を戻そうとして転倒。
	看護師が排尿意図により事故が起きたと状況から判断している場合。	尿意により起き上がったと考えられる。
フラグを除外する	看護師が、患者の行動について排尿意図によるものであるか不明であると明言している場合。	「トイレに行きたかった」と話すが、実際は不明。
	事故後に排尿行為を行なっている場合	事故の後トイレまで歩行されていることや状況から考えて自ら床へ降りたのではないと思われる
	トイレから戻る際の寄道による事故。	トイレから戻ろうとして転倒。
	看護師が、排尿関連の付き添いが終わったと判断した後のイベント。	排尿介助後、ベッドに戻るのを確認したがその後洗面所で転倒
	普段の行動に関する表現。	普段はトイレ後ナースコールがあるが、
	主語が異なる行動。	他患者のトイレ介助のため訪室すると、

ルール

採用語	Start語	トイレ	トイレ ポータブル 用足し 尿器 便座 排尿 排便 排泄 失禁 尿意	プロジェクト開始時に追加
		トイレ	pトイレ Pトイレ p-トイレ P-トイレ pWC pwc pWc PWC Pwc PWc p-WC p-wc p-Wc P-WC P-wc P-Wc 排尿する 排便する 排泄する お手洗い 御手洗 御手洗い 便意 おしっこ オシッコ 大便 うんこ 便 しっこ 便所 尿失禁 お花畑 女子トイレ イレ	
	その他	汚染 オムツ		
	End語	移動を示す	行く いく 向かう むかう 着く つく 行う 寄る 探す 帰る	
		動作を示す	着座する する ずり落ちる たちくらむ 上げる 出る 座り込む ふらつく 立ち上がる 滑る 排尿する 失禁 握る 返答する 移乗する	
本人の意思		する 思う つもり たい しよう ため もよおす 催す 感じる 不快		
他者が確認できる状況		～を(が)見る みる 訴える 訴え 見られる ある 推測する		
その他	自己 こぼれる 漏れる 濡れる			
除外語	否定	ず ない なし 無い 無し		
	患者本人ではない行動	同室者 他患者		
	普段/対策を表す言葉	今後 際は たびに 度に ときは 時は 以前 普段 必ず 忘れる 前 問題ない 回も 多い 夜間 朝方 説明する ごとに 集中する 頻回		
	その他	血液 流血 ふた		
			チューニング過程で追加	

図3 本研究で用いた因子定義(上)とルール(下)の詳細に関する概要図

3. 研究結果

Watson Explorer を用いて、「排尿意図あり」因子の抽出を行った。今回の分析手法で排尿意図に起因するとフラグ付けされた事故は全体の 41%であった(図4)。認知症患者が夜間(21-6 時)に排尿意図により転倒する割合が高く、重点的なケアが必要であることがわかった(表1)。具体的には、全 3439 レコードに対し、1425 レコード(約 41%)が排尿意図に起因していた。この結果は、排尿意図の予兆検知や、排尿のための移動の補助を行うことで、最大 41%の転倒転落事故を防止することが可能だと考えられる。

さらに、認知症有無別に排尿意図による事故件数を集計したところ、排尿意図による事故の件数は認知症患者に多く、割合は非認知症患者が高いことが明らかとなった(図5、図6)。また、性別ごとに集計した排尿意図による事故の割合を、認知症・それ以外で集計したところ、認知症の有無を問わず、性別間の排尿意図による事故の割合の差異は、ほとんどないことがわかった。

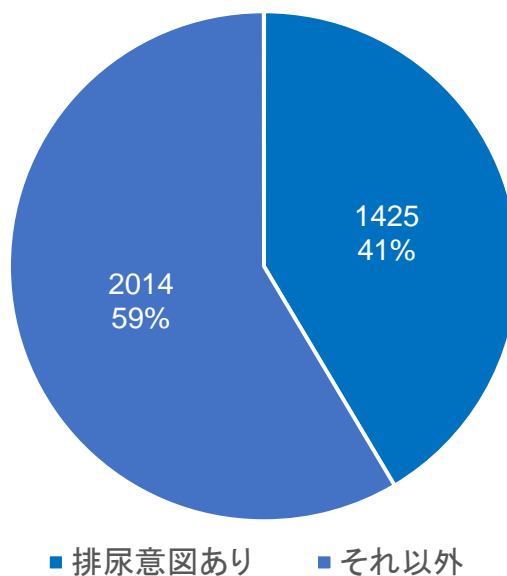


図4 転倒転落事故に占める排尿意図の割合

表1. 発生時間帯別に集計した排尿意図による転倒転落数.

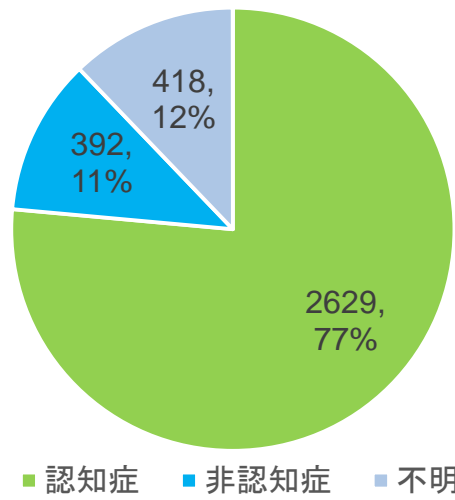


図5 認知症有無による排尿意図の転倒割合

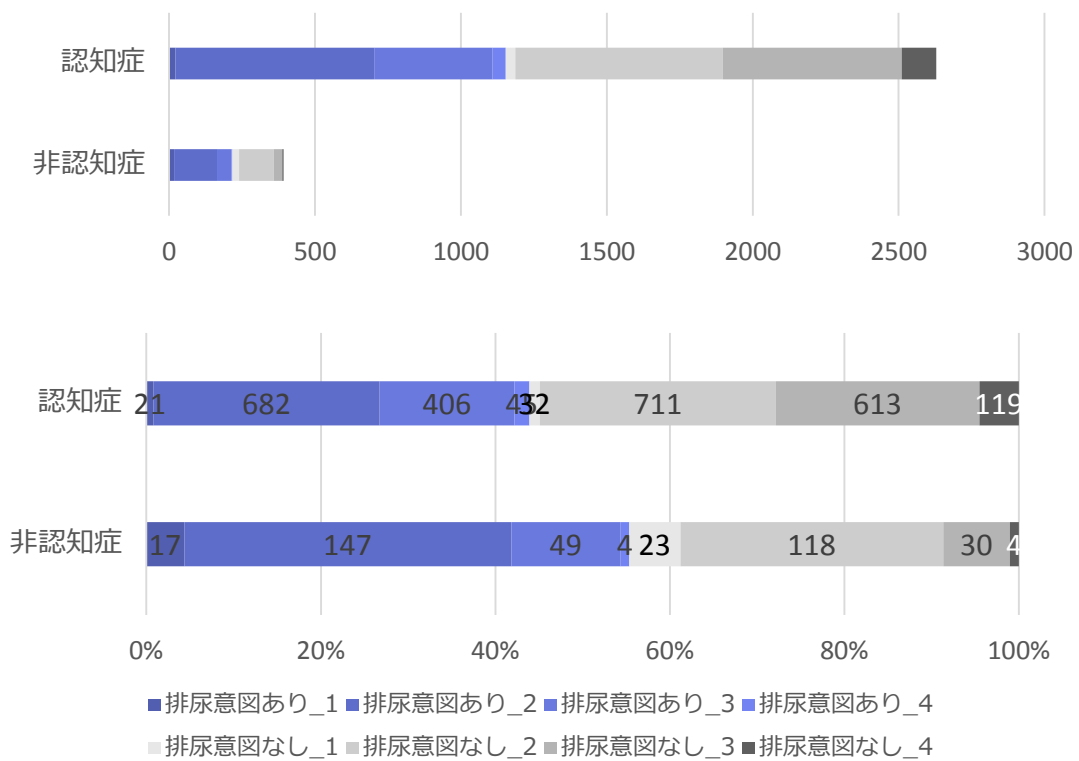


図6. 認知症有無別排尿意図による事故の割合.

各棒グラフにおいて、排尿意図ありを色付き、排尿意図なしを白で示した。また、各項目について患者の運動機能を4段階で分けて表示している(1が最も良くて、4が最も悪い)。

また、性別ごとに集計した排尿意図による事故の割合を、認知症・それ以外で集計した結果、認知症の有無を問わず、性別間の排尿意図による事故の割合の差異は、ほとんどないことがないことが明らかとなった(図7、図8)。

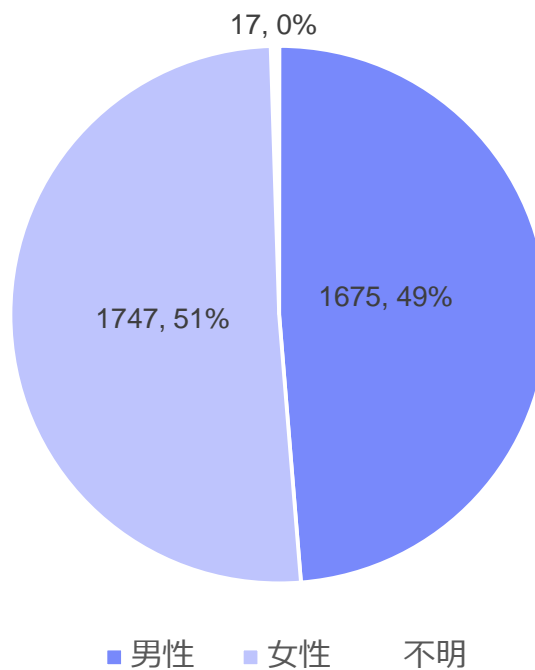


図7. 全体に占める男女比の割合

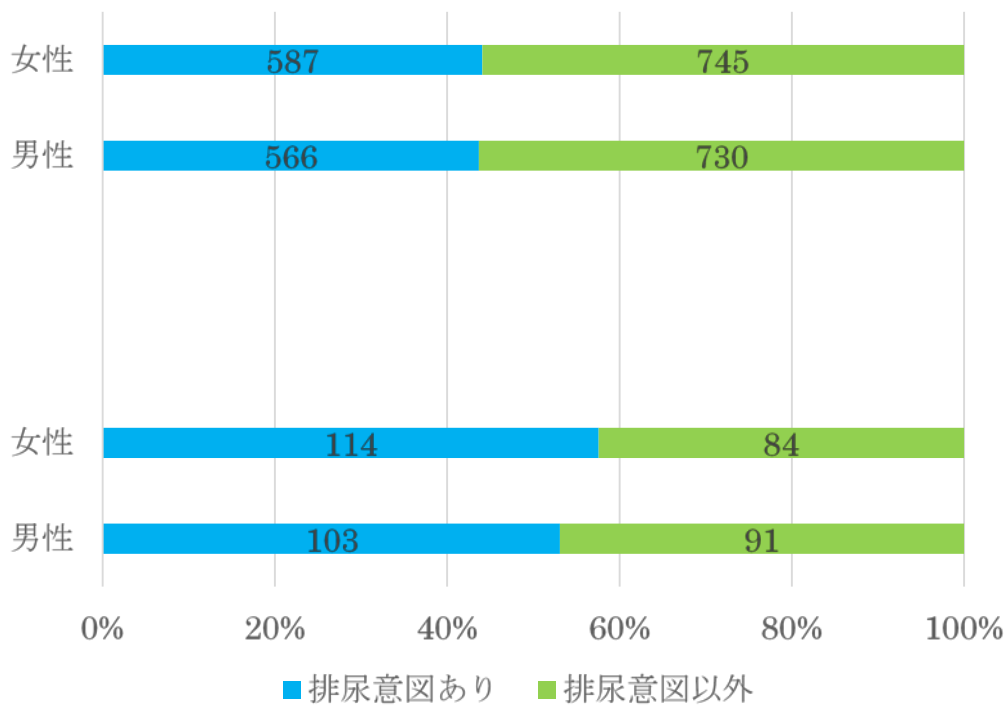


図8. 認知症あり(上)と認知症以外(下)において性別に集計した転倒転落数.

また、年代別に集計した排尿意図による事故の割合を、認知症・それ以外で集計を行った。

その結果、認知症患者においては、年齢が高まるほど、排尿意図に起因する割合が増加する傾向にあることが明らかとなった(図9、図10)。

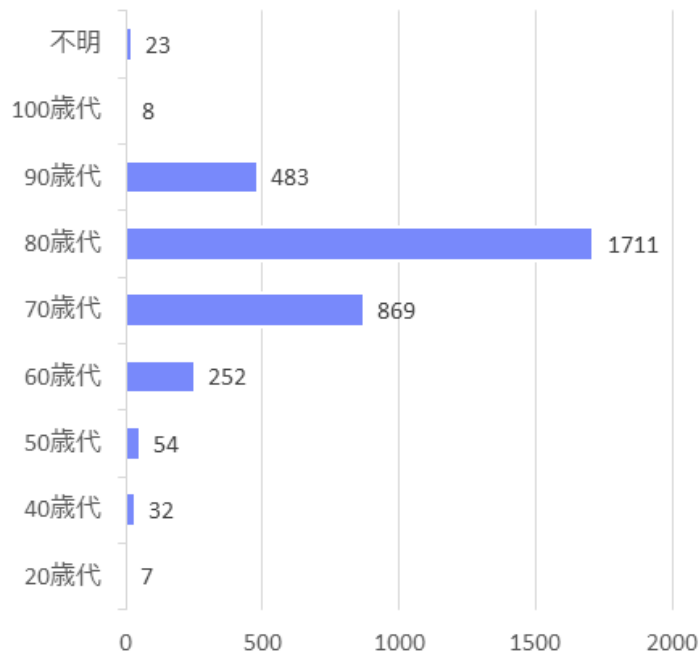


図9. 集計全体に占める年齢別患者数.

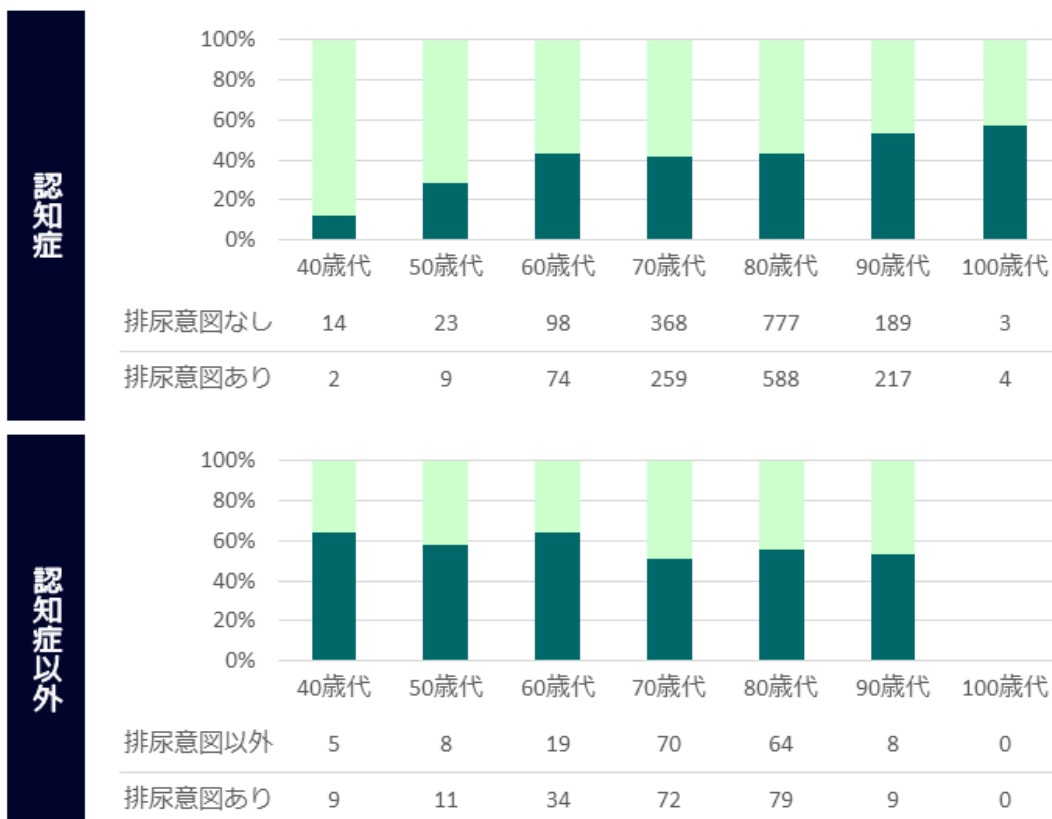


図10. 認知症(上)、認知症以外(下)における年齢別での排尿意図に関連した転倒転落数.

最後に、使用する歩行補助具別に集計した排尿意図による事故の割合を、認知症・それ以外で集計を行った。その結果、補助具が不要な患者の排尿意図による事故割合は極めて低いことがわかった(図11、図12)。

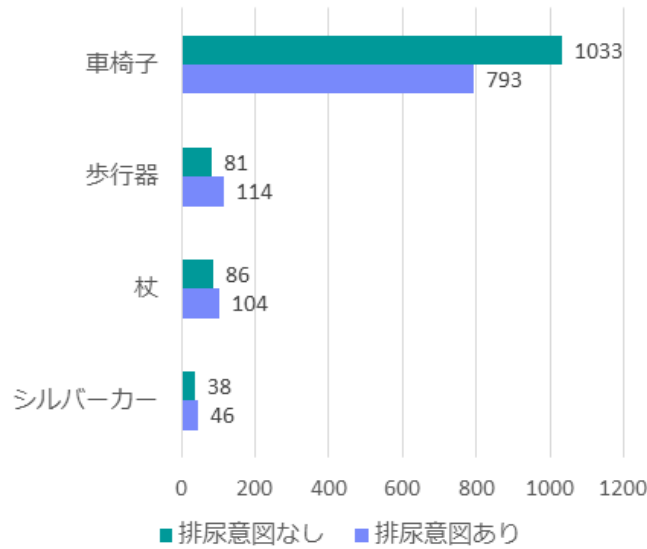


図11. 歩行補助具・排尿意図別事故件数

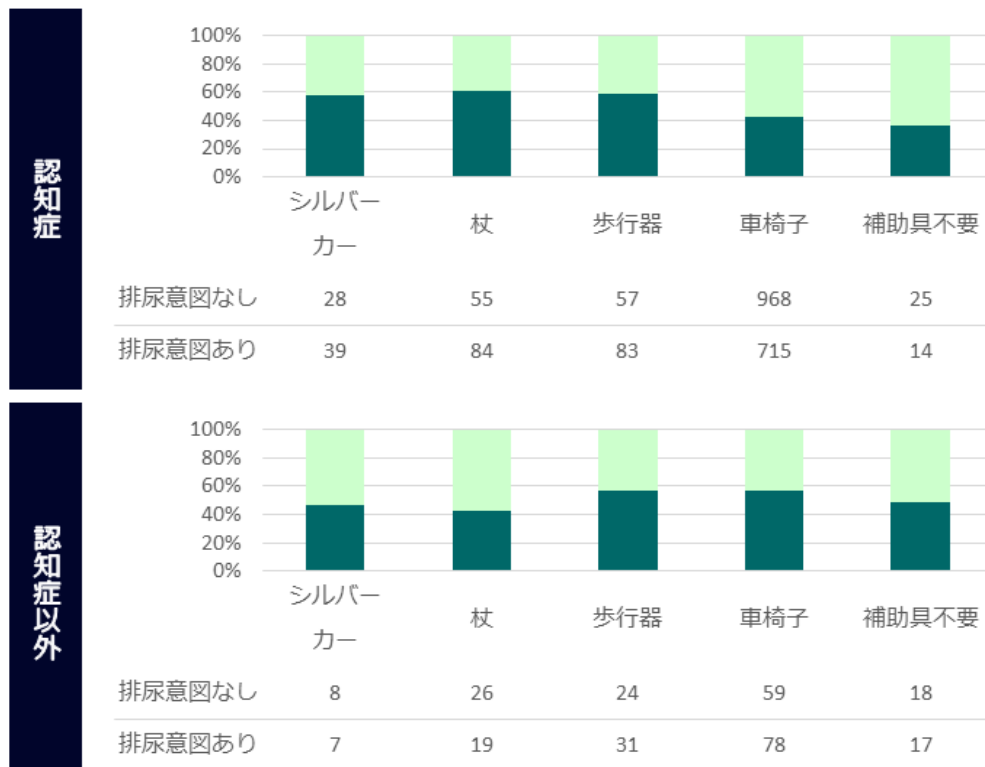


図12. 認知症(上)、認知症以外(下)における歩行補助具・排尿意図別事故件数

4. 研究結果のまとめと考察

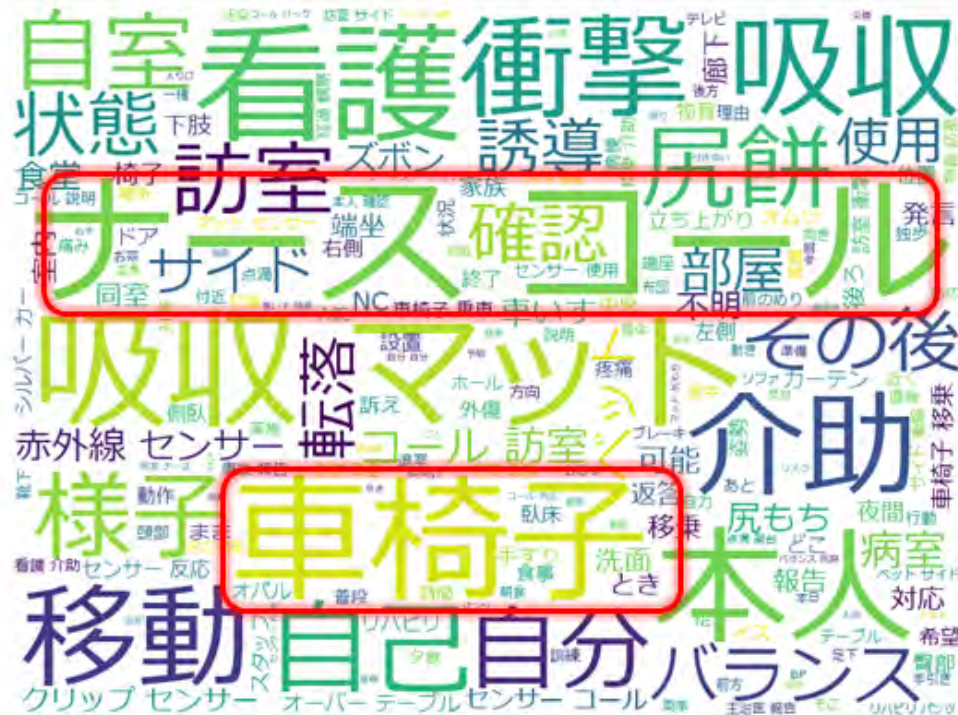
本研究によって Watson Explorer を用いて、当センターでこれまで 3000 件以上を集積してきた認知症入院患者における転倒転落事故報告書の大規模テキストマイニング解析を試みた結果、認知症患者が夜間(21-6 時)に排尿意図により転倒する割合が高く、重点的なケアが必要であることが示唆された。さらに、認知症有無別に排尿意図による事故件数を集計したところ、排尿意図による事故件数は認知症患者に多く、一方で、その割合は非認知症患者の方が高い傾向にあることがわかった。これらの結果は、これまで明らかにされてこなかった認知症患者の行動意図等の内的要因に対する転倒リスクを、統合的に検討できたことを示唆している。

また、これまでの先行研究で示されてきたサンプル数は、高々 1000 件に満たない調査報告であったがゆえに、結果に対する解釈が統一されてこなかった。したがって、本研究では、それら先行研究の約 3 倍に匹敵する、当センターで過去 5 年間にわたって集積した 3000 件以上ものサンプルサイズでの解析に成功し、これまで打ち出せなかった認知症における転倒リスクに対して、一定の見解を提示できたと考える。

本研究で開発したテキストマイニングによる行動意図の抽出は、“尿意”に関連する単語・語句を導出するだけでなく、他の関連因子(歯磨き、ベッド・廊下への移動等)にも適応できる可能性を持っており、今後は、3000 件を超える報告書に対してあらゆる行動・意図の因子解析に適応できるアルゴリズムの開発に繋がっていきたいと考える(参考資料 1-3)。

最終的には、各行動意図に対するリスクを時間帯に分けて評価し、各行動因子と時間帯との関連性も併せて検討することを目標にしたい。

- 集計方法
 - ✓ 「～ようとする」「～したい」といった、意図を表す表現を含む文章に出現する名詞を抽出し、排尿に関連する単語等を除外した後WordCloudで表示
 - ✓ 出現頻度の高い単語をより大きく図示



参考資料1. 排尿意図以外の行動意図に関連する名詞の抽出結果.

行動意図表現と共起する名詞のうち、排尿意図に関連しないものを集計・可視化を行った。効果の高い転倒転落リスク軽減策案として、1.車椅子からの移動補助体制強化 2.ナースコールの紛失防止などが考えられる。

- データの解釈：「車椅子」「ナースコール」という名詞が頻出している。
 - 1. **車椅子**からの移動補助体制強化
 - 2. **ナースコール**の紛失防止
等が、転倒
- 転落事故リスク軽減策として有効あると考えられます。

■ 「ナースコール」を含む文章例：

- ナースコールが手元に無かったので取ろうとした。
- ポータブルトイレに排尿後、ベットの上に置いてあったナースコールを取ろうとして、ずり落ちて尻餅をついた。
→定義に合致(ナースコールを取るための移動)
- X 「トイレに行こうとして転んだ」とナースコールで報告有り。
→**WEXチューニング**が必要(ナースコール≠転倒の原因)

■ 「車椅子」を含む文章例：

- 車椅子からベッドへ**1**人で移乗しようとしたと推測される。
- オパールから車椅子に移動しようとした時に、めまい・膝折れし、後ろに倒れ込んだ。

参考資料2. 排尿意図以外の行動意図の抽出結果に対する解釈、および抽出文章の例

action_surface	action_lemma	record_count	percent_N-T	rc/total_N	rc/total_T
行く	行く いく 行こう いこう	482	-0.0281	345.0 / 2373.0	110.0 / 634.0
食堂	食堂 ホール	262	0.0854	240.0 / 2373.0	10.0 / 634.0
不明	不明 忘れ 覚えていない わからない 分からない	198	0.0275	155.0 / 2373.0	24.0 / 634.0
取る	取る 取るう とろう 取り たかった とりたかった	82	-0.0124	53.0 / 2373.0	22.0 / 634.0
リハビリ	リハビリ	81	-0.0245	43.0 / 2373.0	27.0 / 634.0
戻る	戻る から戻 から帰	75	-0.0145	48.0 / 2373.0	22.0 / 634.0
座る	座る 座ろう すわろう す わる	62	-0.0161	33.0 / 2373.0	19.0 / 634.0
カーテン	カーテン	61	0.0060	48.0 / 2373.0	9.0 / 634.0
呼ぶ	呼ぶ 呼ぼう よぼう	40	-0.0080	26.0 / 2373.0	12.0 / 634.0
更衣	更衣 着る 着よう 着たい 着替え 服を	32	-0.0089	20.0 / 2373.0	11.0 / 634.0
戻す	戻す 戻そう 戻した もど す もどそう	31	-0.0053	21.0 / 2373.0	9.0 / 634.0
テレビ	テレビ リモコン 本を	30	-0.0002	22.0 / 2373.0	6.0 / 634.0
拾う	拾う 拾おう ひろう ひろ おう	29	-0.0134	13.0 / 2373.0	12.0 / 634.0
しまう	しまう しまおう	28	-0.0002	22.0 / 2373.0	6.0 / 634.0
家に帰る	家に帰る 家に帰ろう 家に 帰りたい 帰宅 家に戻	26	0.0073	21.0 / 2373.0	1.0 / 634.0
シャワー	シャワー 風呂 浴室 脱衣 所	22	-0.0171	8.0 / 2373.0	13.0 / 634.0
歯磨き	歯磨き 歯を磨	12	0.0006	9.0 / 2373.0	2.0 / 634.0

参考資料3. 排尿意図以外の行動意図候補の単語例.

排尿意図以外の行動意図候補の単語が記載されたレコードの頻度を認知症・非認知症ごとに集計した。これらの解析を応用することで、「行く」「拾う」といった一般動詞の述語を分類できる可能性がある。

平成 29 年度 老人保健健康増進事業の報告書

事業名: 認知症の診断および介護に対するAIおよびIoTの活用に関する探索的調査研究

国立研究開発法人国立長寿医療研究センター

メディカルゲノムセンター 尾崎浩一 新飯田俊平

研究の目的

本実施事業は、認知症患者における先天的生命情報(全ゲノム情報や RNA 発現情報等)、および後天的生命情報(臨床データ、脳・生体情報等)を組み合わせることで、機械学習やAIを用いた学習による認知症判別モデルを開発し、認知症患者における生活支援機器を創出することを目的として以下の 2 項目について検討、実施したので報告する。

1. 全ゲノム関連解析の施行、解析及び全ゲノム、全 RNA 配列解析の準備と検討

近年、様々な認知症治療薬の臨床治験が進められているが未だ成果が上がっていないのが現状であり、根本的な疾患の原因を突き止めてエビデンスに基づく創薬やドラッグリポジショニング、早期予知による発症の阻止を目指すことが必要である。認知症は環境と遺伝因子が複雑に関与して発症することが知られているが、近年の疫学研究により遺伝因子の重要性が強く浮かび上がってきている(図 1)。このような背景のもと、これまでに欧米において認知症のゲノムワイド関連解析(GWAS)が大規模に施行され、疾患関連座位群が同定されているが、日本人において再現されたのはわずか数座位にとどまっており、ほぼ未解明であると言っても過言ではない。このギャップは欧米人と日本人のゲノム構造の違いに依存すると考えられ、日本人ゲノム構造に特化した解析を進めることが喫緊の課題であり、さらに臨床情報等も加味してその機能的な側面を解明することがエビデンスに基づく診断・治療法開発に重要である。本事業ではゲノム情報、RNA配列情報等のオミックスデータと臨床情報を統合することによる機械学習やAIを用いた学習による認知症判別モデルを開発(図 2、3)するための第一段階として、国立長寿医療研究セン

ターバイオバンクが保有するDNAサンプルを用い、日本人に特化したジャポニカアレイによる一塩基多型 (SNP) の全ジェノタイピングデータおよびバフィーコートによる全RNA配列データの取得を行ってきた。現在までに約 7,100 例 (アルツハイマー病; AD 2,391 例、認知機能正常群; Control 3,346 例、軽度認知障害; MCI 601 例、血管性認知症; VaD 88 例、前頭側頭型認知症; FTD 25 例、レビー小体認知症; DLB 128 例、その他 553rei) の全ゲノムジェノタイピングデータを取得している (図 4)。ここで得られたジェノタイプデータから、試験的に AD 2,250 例、コントロール 1,848 例を用いて、1000 ゲノムデータ (www.internationalgenome.org/) を用いたインピュテーションを施行し、GWASを施行したところ、既報通り *APOE* 座位において非常に有意な AD との関連が認められると同時に (図 5)、集団階層化の指標であるラムダ GC が 1.00 となり、集団の階層化がほとんど存在しないことが示唆された (図 6)。また、本試験的 GWAS においては新規の GWAS 有意性 ($P = 5 \times 10^{-8}$ 未満) を示す染色体領域は見られなかったが、示唆的な有意性 ($P = 10^{-5}$ 未満) を示す座位が第 3、5、8、10、13、15、18 番染色体などに認められ (図 5)、これらの中には低頻度かつ東アジア人にしかアレル頻度が認められない SNP も存在していた。今後サンプル数をさらに増加すると共に、日本人 3,500 人から得られたインピュテーションのリファレンスパネルを用いることによりさらに有意な新規座位が得られるものと考えられる。

Table 2. Parameter Estimates (With 95% Confidence Intervals) From Model Fitting Results for Alzheimer Disease: Full Model Without Adjusting for Age and Full and Reduced Models Adjusting for Age*

Group	遺伝率	共有環境	非共有環境	r_g for Unlike-Sex Pairs	Threshold†
	A*	C*	E*		
Full Model Without Adjusting for Age					
Men	0.59 (0.00-0.89)	0.21 (0.00-0.72)	0.20 (0.07-0.46)		1.96
Women	0.37 (0.11-0.63)	0.53 (0.29-0.73)	0.10 (0.05-0.20)		1.47
Unlike sex				0.50	2.14
Full Model Adjusting for Age					
	A*	C*	E*	r_g for Unlike-Sex Pairs	β_{age}
Men	0.58 (0.00-0.86)	0.07 (0.00-0.59)	0.36 (0.14-0.69)		-0.08
Women	0.45 (0.02-0.88)	0.36 (0.00-0.71)	0.19 (0.09-0.35)	0.50	-0.09
Reduced Models Adjusting for Age					
Men and women equated	<u>0.58 (0.19-0.87)</u>	<u>0.19 (0.00-0.49)</u>	<u>0.23 (0.13-0.37)</u>	0.50	-0.09
Men and women equated, C parameter dropped	<u>0.79 (0.67-0.88)</u>	NA	<u>0.21 (0.12-0.33)</u>	0.50	-0.09

Abbreviations: A, additive genetic influences; β_{age} , regression coefficient for age; C, shared environmental influences; E, nonshared environmental influences; NA, not applicable; r_g , genetic correlation.

*All values are parameter estimate (confidence interval).

†Threshold is the z score corresponding to prevalence.

(Mortimer et al. Arch Gen Psychiatry. 2006)

遺伝率 58%~79%

図1 アルツハイマー病の遺伝率

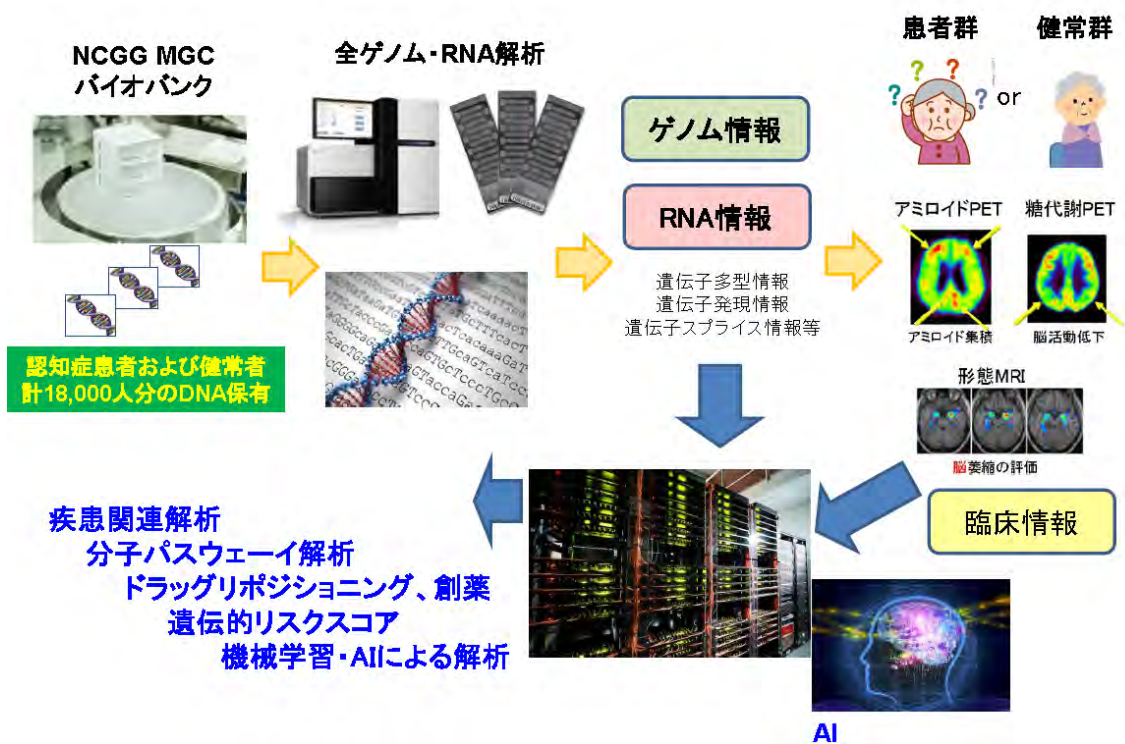


図2 大規模全ゲノム・RNA データの収集と解析

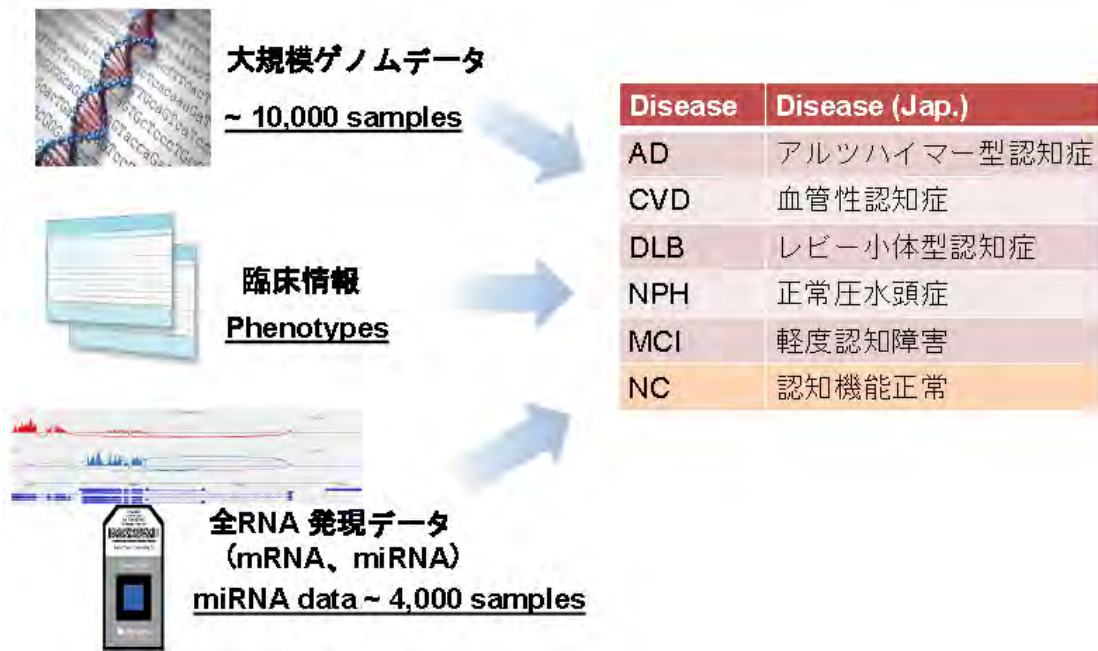


図3 大規模データ統合による疾患予測モデルの構築

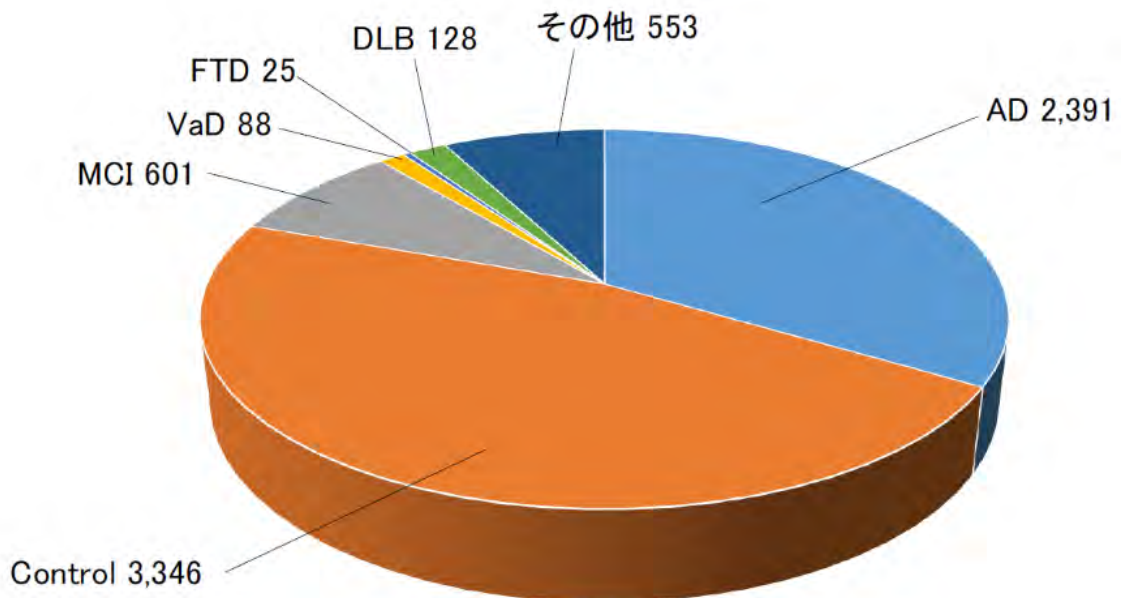


図4 ジャポニカアレイによる全ゲノムジェノタイピングの分布

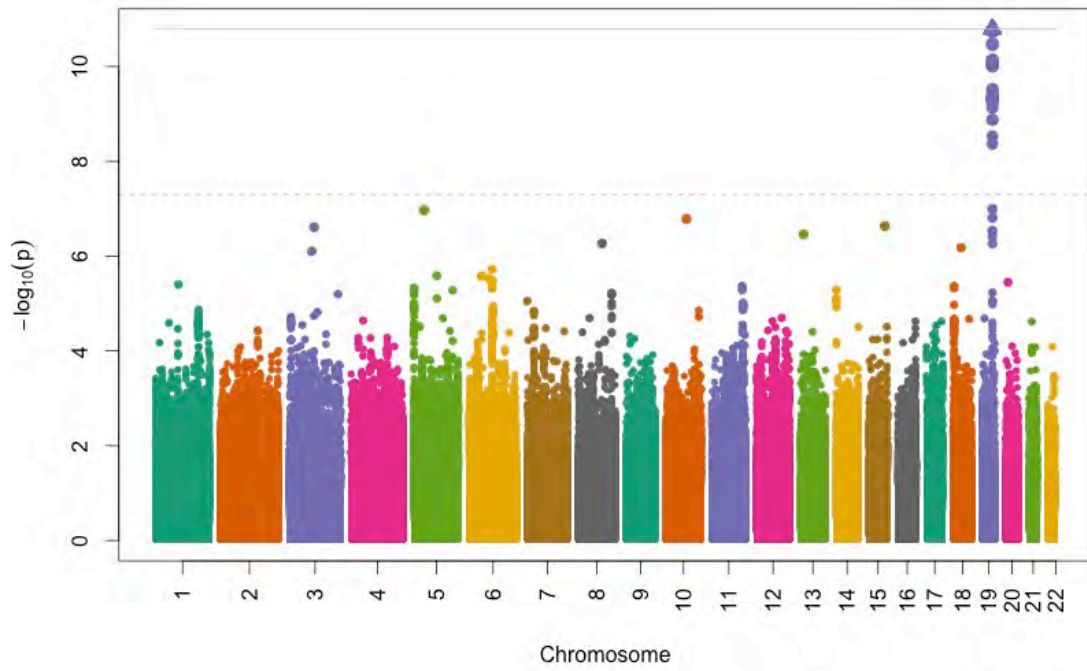


図 5 試験的 GWAS のマンハッタンプロット

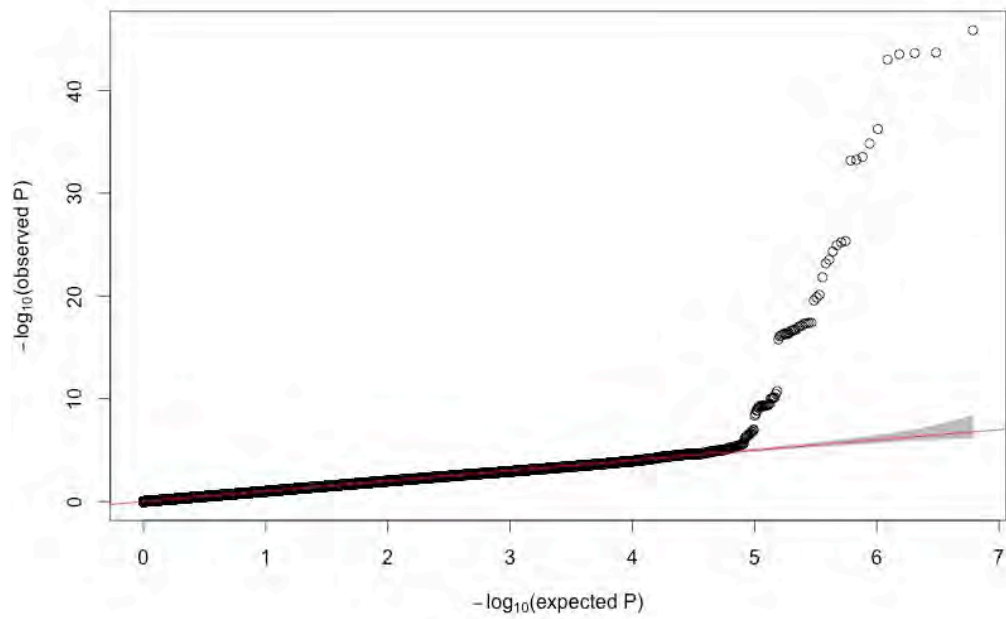


図 6 試験的 GWAS の Q-Q プロット

また、全RNA配列の取得については、約 200 例について高品質なRNAを取得し、全RNA配列解析用のライブラリ作製を進めてきた(図 6、7)。全RNAの質の指標であるRNA integrity number (RIN) は概ね 7.0 以上が得られており、RNAライブラリの質も良好であった。今後は、全ジェノタイプングデータを 10,000 例まで増やすとともに、全RNAについては次世代シーケンサーにより全RNA配列データを取得する予定であり、その後これらの統合解析から機械学習、AIによる認知症予知予測法の確立を目指すと共に、ドラッグリポジショニングによる既存治療薬の認知症への適応、新規創薬のターゲット探索を進めていく。

Gel Image

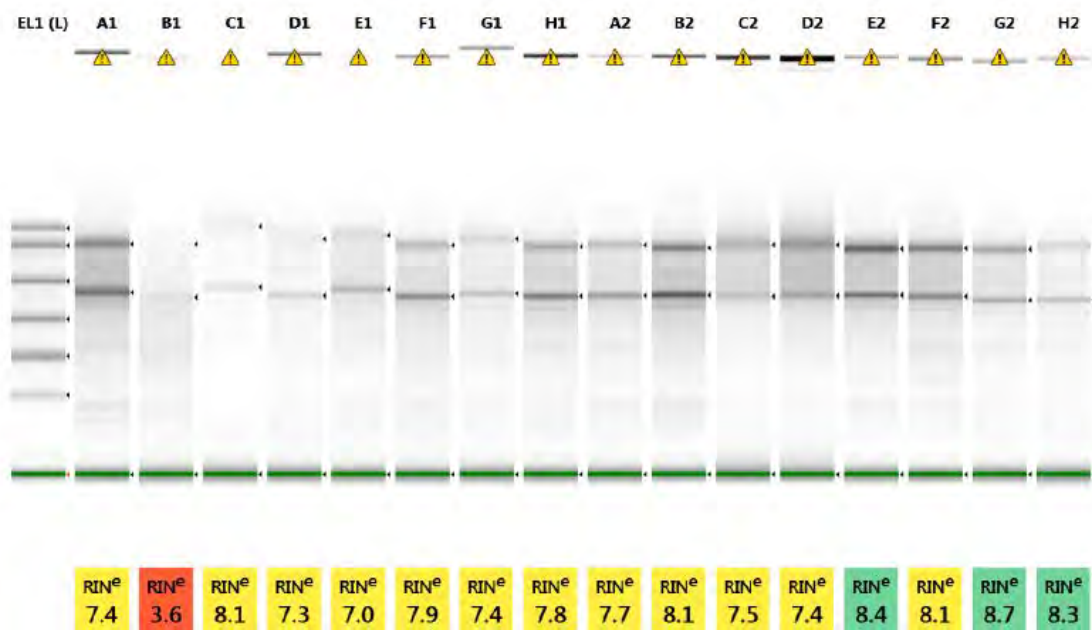


図7 バフィーコートから抽出したRNAのクオリティー 図5 全RNAライブラリのクオリティー

Gel Image

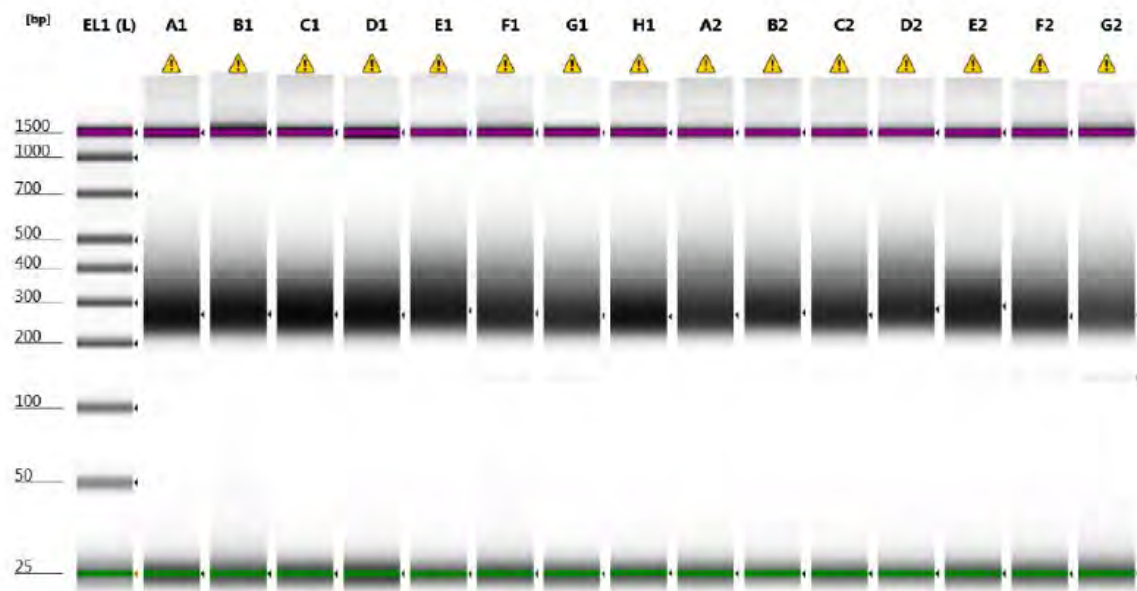


図8 全RNAライブラリのクオリティー

2. miRNA 発現情報と臨床情報の整備及び深層学習による認知症タイプの判別法の検討

主な認知症には、アルツハイマー病、脳血管性認知症、レビー小体型認知症、前頭側頭葉型認知症がある。これらの判別を目的として、国立長寿医療研究センター・メディカルゲノムセンターが参画するAMED 体液中マイクロ RNA 測定技術基盤開発プロジェクトで取得した 2,000 名を超える認知症患者および健常高齢者の血液中の約 2,500 種のマイクロ RNA (miRNA) 発現量の解析データと臨床データから機械学習を行うことで、アルツハイマー病 (AD)、脳血管性認知症 (VaD)、レビー小体型認知症 (DLB) 等の、疾患タイプを判別する予測モデル開発の検討を進めた。本件研究では、miRNA 発現量の主成分 (principal component、PC) に着目し、その違いが各種認知症疾患を分別できる可能性があるという仮説に基づき (図 9)、検討を行った。PC については、あらかじめデータを分類した教師あり学習 (Supervised PC) に着目し、疾患判別モデルの構築を試みた。図 10 に本解析に用いた疾患およびコントロールサンプルの内訳、特徴を示した。サンプルとしてほぼ半数にランダムに分けた探索サンプル (Discovery set) と再検証サンプル (Validation set) を用いた。miRNA を抽出するために統計量として T 値を用い、抽出した miRNA 群の主成分分析により次元を圧縮した。その後、PC スコアを基準とした予測モデルの構築を進め、クロスバリデーションを 10 回繰り返す機械学習によって疾患予測モデルを確立した。さらに再検証サンプルセットを用いた検証を試みた。

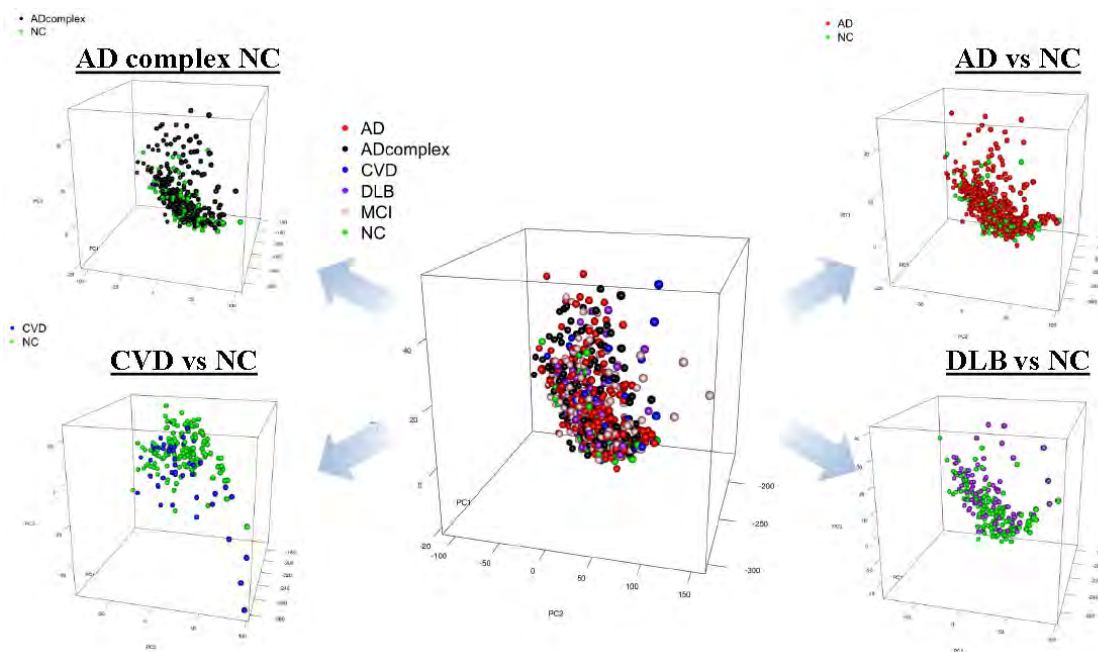


図9 各疾患でのmiRNA発現における主成分を用いた疾患—コントロールの分離

Disease	#Sample	Discovery set	Validation set
AD (Alzheimer's disease)	1,021	511	510
VaD (Vascular Dementia)	91	46	45
DLB (Dementia with Lewy Bodies)	169	85	84
MCI	612	306	306
NC	288	144	144
Total	2,181	1,092	1,089

Disease	Discovery set (Disease +NC)			Validation set (Disease +NC)		
	Age	Sex (Male)	APOE	Age	Sex (Male)	APOE
AD	77.6	0.33	0.46	77.6	0.37	0.40
CVD	74.6	0.48	0.27	74.6	0.48	0.21
DLB	73.4	0.53	0.25	73.5	0.56	0.16

図10 疾患およびコントロールサンプルの内訳、特徴

結果として、AD、VaD、DLBとコントロールを判別する予測モデルの構築を行ったところ、AD、VaD、DLBにおける探索セットではAUC(クロスバリデーション)がそれぞれ0.87程度となり、再検証セットでも同様のAUCが得られ好成績となった。また、これらの正確度(Accuracy)、感度(Sensitivity)、特異度(Specificity)も非常に高い数値を示した。

現在、本モデルは前向き患者コホートによる認知症発症予測における精度の確認を進めている。また、今後は1.を進めることにより得られる日本人特有の全ゲノムジェノタイピングデータ、全RNA配列解析データおよび画像データも含めた詳細な臨床データを使用するビックデータを用いて統合解析を行い、深層学習およびAIによりさらに正確な疾患診断法等の研究開発に着手する。

参考資料

日本人類遺伝学会第62回大会 発表資料(平成29年11月17日)

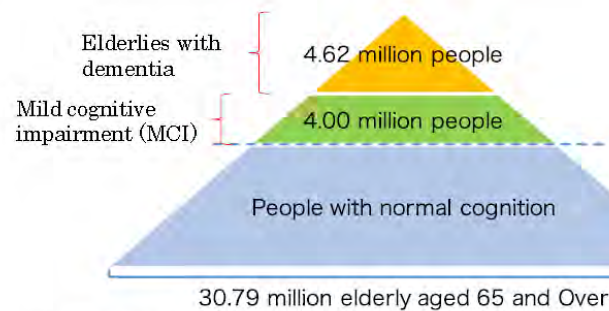
日本人における孤発性アルツハイマー病のゲノムワイド関連解析

Genome wide association study for late-onset Alzheimer's Disease in a Japanese population

Risa Mitsumori¹, Yuya Asanomi¹, Daichi Shigemizu^{1,2,3}, Yuki Nagata¹, Taiki Mori¹, Shintaro Akiyama¹, Shumpei Niida⁴, Kouichi Ozaki¹

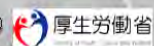
1) Laboratory for Medical Genome Sciences, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Medical Genome Center, NCGG, Obu, Japan; 5) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Elderlies with Dementia



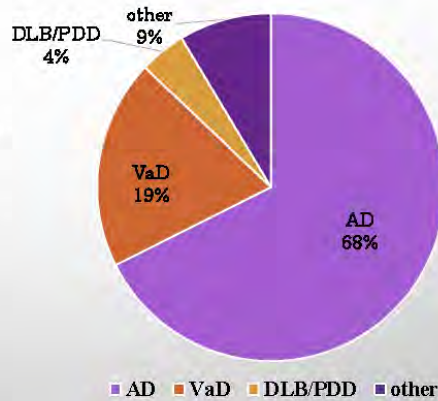
For 65-year-old or older nationwide, dementia prevalence rate was estimated 15%, and it was estimated that the MCI prevalence were 13%.

Ministry of Health, Labour and Welfare (2012)



Types of Dementia

- Alzheimer's disease (AD)
- Vascular dementia (VaD)
- Dementia with Lewy bodies (DLB)
- Mixed dementia
- Parkinson's disease (PD)
- Frontotemporal dementia (FTD)
- Creutzfeldt-Jakob disease
- Normal pressure hydrocephalus
-
-
-
- etc.



T Asada
Dementia prevalence rate within urban area and correspond to life function disorder (2013)

Alzheimer's disease (AD)

The Progression of AD

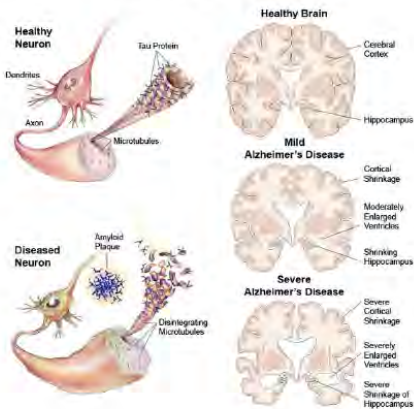


Illustration by Bob Merritt, provided courtesy of the BrightFocus Foundation



Types of AD

Familial AD (FAD)

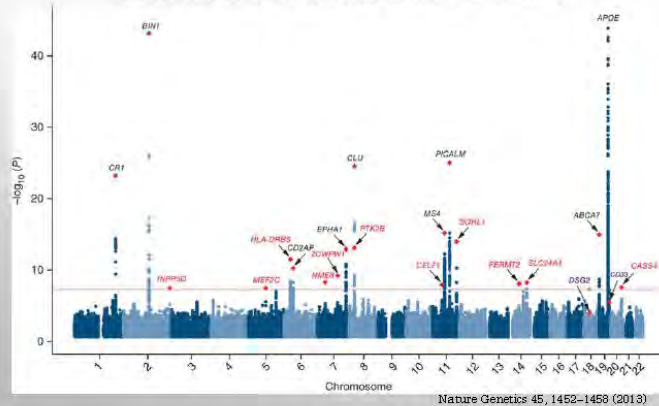
- An age of onset is early (under 65 years old)
- High heritable (penetrance > 95%)
- Causative genes : *APP, PSEN1, PSEN2*

Late-Onset AD (LOAD)

- The most common form of dementia
- Sporadic
- Heritability : 58~79%
- Risk gene : *APOE ε4/ε4* (OR=11.6)
- Causative genes : *ε3/ε4 & ε2/ε4* (OR=3.2)
- *ε3/ε3* (OR=1.0)
- *ε2/ε3* (OR=0.6)
- (Heritability ~10%)

Risk variants for LOAD

GWAS for LOAD in Caucasian population

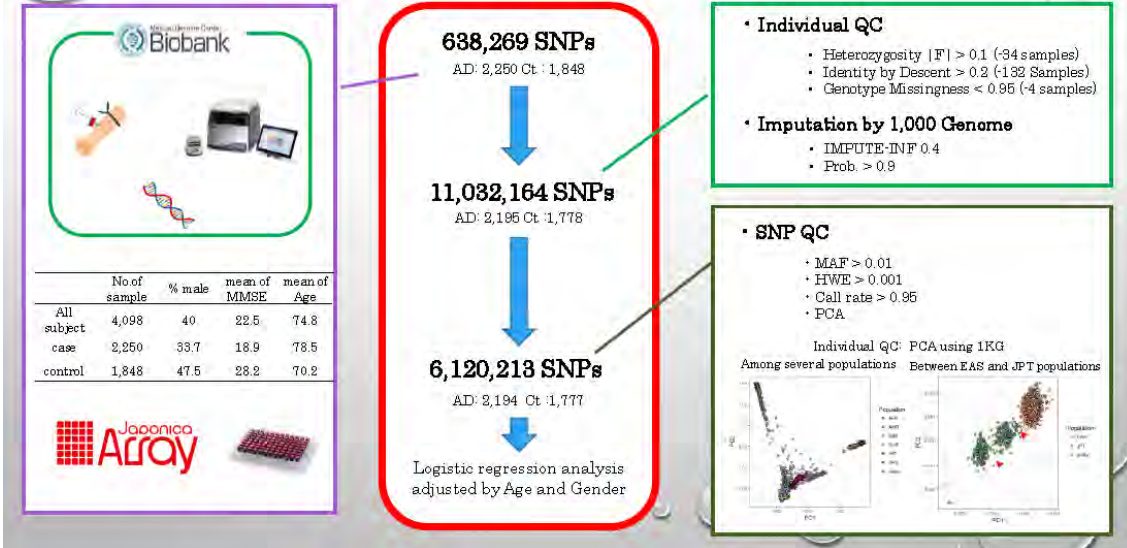


Three variants of known LOAD loci in Caucasian population (*SORL1*, *PICAM* and *BIN1*) were replicated in Japanese.

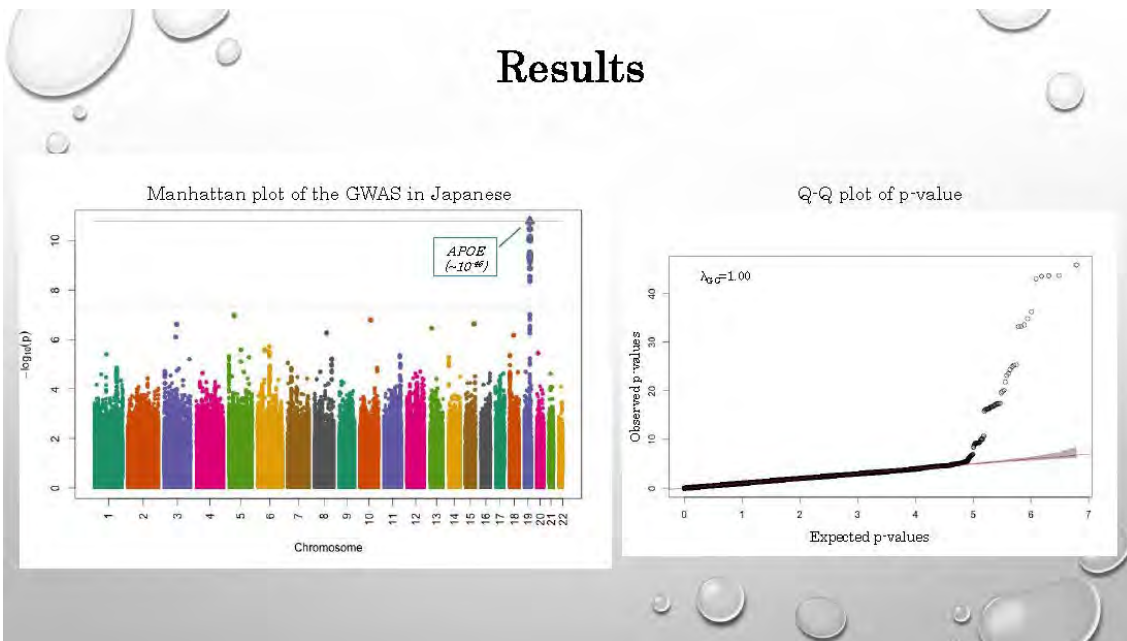
Aim

- As a part of “**Program for an Integrated Database of Clinical and Genomic Information for AMED**”, we aim to do genotype with 10,000 samples using Japonica Array which includes Japanese specific SNPs followed by dementia GWAS and data storage for integrated database.
- We have tentatively conducted GWAS of Alzheimer’s disease in Japanese population with approximately 2,300 cases and 1,900 controls.
- We also attempted to replicate 24 known loci identified in Caucasian population for LOAD in Japanese individuals.

Materials and Methods



Results



Replication for 24 known loci in Japanese individuals

ID	Gene	Chr	OR	P
rs4147929	<i>ABCA7</i>	19	1.181	0.00174
rs616338	<i>AB19</i>	17	NA	NA
rs6733839	<i>BINI</i>	2	1.074	0.1888
rs7274581	<i>CASS4</i>	20	NA	NA
rs1094836	<i>CD2AP</i>	6	NA	NA
rs3866444	<i>CD33</i>	19	1.024	0.7101
rs10838725	<i>CELF1</i>	11	0.9719	0.6418
rs9381896	<i>CLU</i>	8	0.8493	0.005223
rs6666401	<i>CR1</i>	1	1.197	0.1654
rs8093731	<i>DSG2</i>	18	0.9203	0.5154
rs11771145	<i>EPHA1</i>	7	0.9912	0.8654
rs17125944	<i>FERMT2</i>	14	1.022	0.7158
rs9271192	<i>HLA-DRB5- HLA-DRB1</i>	6	NA	NA
rs35349669	<i>INPP5D</i>	2	0.6968	0.07303
rs190982	<i>MEF2C</i>	5	1.063	0.3767
rs983392	<i>MS4A6A</i>	11	1.044	0.6979
rs2718058	<i>NME8</i>	7	1.07	0.3135
rs10792832	<i>PICALM</i>	11	0.8297	0.0004957
rs72824905	<i>PLCG2</i>	16	NA	NA
rs28834970	<i>PTK2B</i>	8	1.045	0.4965
rs10498633	<i>SLC24A4RIN3</i>	14	0.9145	0.4411
rs11218343	<i>SORL1</i>	11	0.8368	0.00133
rs143332484	<i>TREM2</i>	6	NA	NA
rs1476679	<i>ZCWPW1</i>	7	0.9636	0.5092

Summary

We replicated the *APOE* locus ($P = \sim 10^{-46}$) and identified several new candidate susceptible loci for LOAD with suggestive significance. We also found convincing associations for 4 known AD loci, *CLU*, *PICALM*, *ABCA7*, *SORL1*, discovered in Caucasian population, so far.

Future

We are going to do genotype 10,000 samples using Japonica array followed by dementia GWAS to find novel loci for LOAD and public data storage.