

高齢健常者のエクソーム解析

(疾患責任遺伝子同定のためのコントロールとして)

 株式会社DNAチップ研究所

所在地 〒230-0045
神奈川県横浜市鶴見区末広町1-1-43
URL <http://www.dna-chip.co.jp/>

Tel (代表) 045 (500) 5211
(営業) 045 (500) 5225
(技術サポートセンター) 045 (500) 5218
E-mail dnachip-support@dna-chip.co.jp



要旨

次世代シーケンサーから得られたゲノム配列は、ヒトリファレンス配列と比較を行い、塩基の違いを遺伝子変異として検出するのであるが、人種や多型の違いにより健常者であっても、多数の遺伝子変異が検出される。このことから、我々は、今後増えるであろう日本人の疾患患者の変異データと比較を行うために、独自の健常者データを取得する事にした。

まずは、70代の男性4名（過去1年の入院歴なし、大病の既往歴なし）について、エクソーム解析を行った。タンパク質にインパクトを与える変異およびdbSNP JPTと1,000 Genomesの東アジア人アレル頻度1%以下で変異を絞り込んだところ、1名あたり約1,000の変異が残った。また、2名以上で見られる共通変異については600変異が検出され、このうちアミノ酸置換変異でのSIFT予測で”DAMAGING”な物は49変異見つかった。

高齢健常者2名以上で見られる変異は、疾患責任遺伝子の可能性は低い。このためこれらのデータを患者と比較することで、疾患責任遺伝子の同定に役立つことが期待される。

動機

健康な高齢者で見られる変異は、疾患責任遺伝子の可能性は低いと考えられるため、これらのデータを整備・蓄積し、遺伝病患者と比較することで、疾患責任遺伝子の同定・フィルタリングに役立たせたい。

被験者について

- ・ 日本人70代男性4名
 - 過去1年の入院歴なし、大病の既往歴なし
 - 問診票および臨床情報の取得を行っている

※ プロジェクトについては、倫理審査委員会の了承済み

実験条件

- ・ サンプル
 - ヒト 4検体 (D01, D02, D03, D04)
- ・ エクソーム濃縮キット
 - Agilent社製 SureSelect all exon V5+UTRs kit
- ・ シーケンサー
 - illumina社製 HiSeq2000
 - 4検体マルチインデックス
 - Paired-end 100bpで1レーン使用

解析フロー

1. 生のリード情報からのアダプタ配列やクオリティの悪い塩基のトリミング (除去)

2. BWAによる、リード情報のリファレンスゲノムへのアライメント (マッピング)

3. Picardによる、Duplicate リード (全く同じゲノム位置の同じ配列のリード) の除去

4. GATKによる、1000 GenomesのInDel情報に基づいた既知InDel周辺の再アライメントおよび塩基スコアのキャリブレーション

5. GATKによる、リードカバレッジの計算

6. GATKによる、エクソームターゲット領域上のSNV (SNP), InDel変異の検出およびクオリティの悪い変異のフィルタリング

7. snpEffによる、SNV (SNP), InDel変異の効果予測 (Effect解析) とアノテーションの付与

8. dbSNPの日本人アレル頻度情報と1000 Genomesの東アジア人 (日本人+中国人) アレル頻度情報等の付与

9. タンパク質にインパクトを与える変異とアレル頻度情報による変異の絞り込み

10. SIFTによるアミノ酸置換変異とフレームシフト変異のタンパク質への影響予測

リードカバレッジ集計結果

■ リード数集計結果

リードのトリミングおよびDuplicateリード（全く同じゲノム位置の同じ配列のリード）を除去した後のリード数（ペア）の集計結果。

サンプル名	オリジナルペアリード数	トリミング後ペアリード数	Duplicateリード除去後ペアリード数
D01	100,353,174	79,407,838	73,413,290
D02	129,893,720	103,197,972	92,325,715
D03	86,790,051	68,830,832	63,725,707
D04	64,665,415	51,659,678	48,333,999

■ リードカバレッジ集計結果

GATKのDepthOfCoverageを用いて、エクソームターゲット領域（SureSelect Human All Exon V5+UTRs: 117Mb）のリードカバレッジの計算および集計結果。

サンプル名	総塩基数	平均値	25 percentile (第1四分位数[Q1])	中央値 (Q2)	75 percentile (第3四分位数[Q3])	10x以上 (%)	25x以上 (%)	50x以上 (%)	75x以上 (%)	100x以上 (%)
D01	9,135,133,245	78.20	105	62	31	94.6	80.3	58.9	40.6	26.9
D02	11,536,882,163	98.76	132	79	40	96.5	85.4	67.8	51.8	38.1
D03	7,997,351,837	68.46	92	54	27	93.1	76.9	53.2	34.2	21.3
D04	6,087,085,271	52.11	70	41	21	89.6	69.0	40.5	22.0	12.2
Total	34,756,452,516									

※ エクソームターゲット領域はSureDesignから取得した100bp Padded BEDファイルを使用。

※ SureSelect Human All Exon V5+UTRsのプローブ領域としては75Mb。

変異解析集計結果

変異絞込み

サンプル名	総変異数	フィルタリング後 総変異数	タンパク質にインパクト を与える変異数 ※1	アレル頻度1%以下 の変異数 ※2	SIFTで“Damaging” となった変異数 ※3
D01	129,572	120,454	11,376	1,005	230
D02	131,800	121,892	11,432	1,034	230
D03	128,097	119,632	11,341	961	197
D04	125,946	118,155	11,325	1,014	217
全4検体	217,390	197,707	18,934	2,658	713

1. タンパク質にインパクトを与える変異は、ミスセンス、ナンセンス、フレームシフト、スタートロスト、スプライスサイト、InDel等
2. アレル頻度1%以下の変異は、dbSNPの日本人 (JPT-1411) アレル頻度 ≤ 0.01 かつ 1000 Genomesの東アジア人 (日本人+中国人) アレル頻度 ≤ 0.01
3. SIFTはアミノ酸置換変異とフレームシフト変異のタンパク質への影響予測を行うデータベース

高齢者2名以上で見られたレア変異

タンパク質にインパクトを与え、なおかつアレル頻度1%以下の変異。
以下の488遺伝子中に600変異が検出された。

AC004076.6	ASPN	CCDC83	DAP	FBXO2	ICMT	KRTAP9-8	MLL3	OR7C2	RALY	SH3BGR	TMIE	ZNF384
AC005220.3	ATF5	CCNJ	DCAF7	FLG	IGFBP2	LAMC1	MMP17	OR7E24	RBBP6	SH3TC2	TNFAIP6	ZNF565
AC009113.1	ATG3	CCNYL2	DCHS1	FLJ22184	IGFN1	LARP6	MPRIP	OR8B2	RBM19	SIGLEC8	TNXB	ZNF66P
AC010536.1	ATN1	CD34	DCP1B	FMNL1	IL32	LCA10	MRPS34	OVGP1	RBMXL3	SIRPA	TPRXL	ZNF714
AC073082.1	ATP8A2	CDCP2	DDHD1	FMNL2	INF2	LCE4A	MST1L	PABPC1	REC8	SIRPB1	TPSAB1	ZNF720
AC073343.1	ATXN1	CDKL4	DENND4B	FOXC1	INO80E	LFNG	MUC12	PALM2-AK	REG3A	SIRT1	TPSD1	ZNF778
AC074212.3	ATXN2	CDKN2AIP	DENND4C	FOXO1	INS-IGF2	LIMD1	MUC13	PAPPA-AS	RFX1	SKIDA1	TPTE	ZNF804A
AC090616.2	BAIAP2L2	CELA1	DHCR7	FOXO1	IRF2BPL	LINC00636	MUC19	PARP4	RHBG	SLAIN1	TRAK1	ZNF816
AC134043.1	BCAR1	CELA3A	DIAPH1	FOXF2	IST1	LIPE	MUC2	PC	RNF19B	SLC19A1	TREH	
ACIN1	BCL6B	CENPV	DIXDC1	FOXO6	ITIH5	LOR	MUC20	PCDH12	RNF208	SLC37A4	TRIM16	
ACSM5	BRD8	CEP104	DKFZP686J	FRRS1L	ITIH6	LRRC46	MUC3A	PCL0	RP11-124D	SLC38A10	TRIM49	
ADAMT58	BRI3BP	CES5A	DMKN	GAPVD1	KAT6B	LRRC49	MUC5B	PCSK6	RP11-131H	SLC3A2	TRIM49C	
ADAMTSL1	BSPRY	CHD5	DNAJC9	GLIPR1L2	KCNC3	LTBP4	NACA	PCSK9	RP11-1407	SMARCA2	TRIM5	
ADRA2B	C11orf80	CHRNA3	DNHD1	GOLGA6L2	KCNJ12	LUNAP1L	NANOS1	PDE11A	RP11-17M1	SMPDL3B	TSPYL1	
AFAP1L2	C14orf23	CHST15	DOK3	GOLGA8R	KCNN3	MADCAM1	NCL	PDE4DIP	RP11-181C	SNAPC4	TTC40	
AGAP3	C16orf11	CKAP4	DPF3	GPATCH4	KDM6B	MAFA	NCOR2	PELP1	RP11-210N	SNRPC	TTL3	
AHNAK2	C16orf3	CLONKB	DPP6	GPR112	KIAA0040	MAGEC1	NIPA1	PHF2	RP1-130H1	SP9	UBAP2L	
AKD1	C1orf222	CLEC11A	DTNBP1	GPR137	KIAA0232	MAGEF1	NLRC5	PI4KAP2	RP11-492D	SPAG1	UNC99	
AL358813.2	C1orf86	CLEC18B	DYRK1B	GPRIN1	KIAA0430	MAGI1	NOP16	PINX1	RP1L1	SPATA20	UNC93B1	
AL391005.1	C20orf132	CLIC6	EBLN2	GPRIN2	KIAA2018	MAML2	NOTCH2	PITRM1	RP3-368B9	SPEN	URI	
AL589765.1	C21orf49	CLTCL1	EFCAB2	GPSM1	KIF1A	MAML3	NOTCH2NL	PKD1L3	RP5-874C2	SPI1	USP20	
AL691479.2	C21orf62	CLYBL	EIF3G	GPX1	KISS1	MANEA	NOTCH4	PLA2G6	RPL13A	SPRR3	USP26	
ALG10B	C2orf71	CNDP1	EMG1	GYLTL1B	KLF14	MAP2K3	NOX4	PLEKHG4B	RPL14	SRA1	VEGFA	
ALMS1	C4A	CNPPD1	EOMES	HADHB	KPNA4	MAP3K1	NPIPL2	PODXL	RRN3	SRRD	VILL	
ALPK2	C4orf6	CNTN5	EP400	HAVCR1	KRI1	MAP3K4	NR1H2	POLDIP2	RSP01	SRRM5	VKORC1L1	
ANAPC1	C8orf42	CNTNAP2	EPHA6	HHIPL1	KRT1	MAPK15	NUAK1	POLR1B	RTTN	SSC5D	VPS13B	
ANKLE1	C8orf59	COL6A2	ESPN	HLA-A	KRT10	MCC	NUDT11	POU2F1	RXFP1	STAG2	VPS9D1	
ANKMY1	C9orf174	COL7A1	ESPNL	HLA-DQA1	KRT2	MCM10	NUP153	POU4F2	RYK	STAU1	VSIG10	
ANKRD33B	CABP1	COPZ2	EVI5	HLA-DQB1	KRT4	MED15	NUP210L	PPYR1	S100A16	STRA8	VSIG10L	
ANKRD36	CACNA1B	CREB3L1	EVPL	HLA-DRB5	KRTAP10-7	MEF2A	OBP2A	PREP	SAFB2	SUSD2	WDFY4	
ANKRD65	CACNA2D1	CREB3L2	FAM182B	HNF1A	KRTAP1-3	MEGF11	OGFR	PRKCSH	SAMD1	SYNGR1	WDR66	
ANKS1A	CAMKK2	CRIPAK	FAM186A	HOMER3	KRTAP17-1	MEGF9	OLFM1	PRKDC	SCAF1	SYNJ1	WRN	
AP003026.1	CAPN12	CROCCP2	FAM187A	HRC	KRTAP29-1	MEOX2	OPN1LW	PRR21	SCARF2	SYT7	YEATS2	
AR	CARD10	CSPG4	FAM189A2	HRNR	KRTAP4-3	MEX3A	OR13C2	PRR23A	SCUBE2	TBP	ZDHHC11	
ARD1B	CAST	CTSA	FAM194B	HSH2D	KRTAP4-7	MEX3D	OR2L2	PRSS3	SEC16A	TCEAL4	ZFH3	
ARSD	CCBE1	CYFIP2	FAM205A	HSPBAP1	KRTAP4-8	MGC20647	OR2T12	PRSS48	SELPLG	TCEAL6	ZNF107	
ART5	CCDC120	CYP11B1	FAM22F	HSPBP1	KRTAP4-9	MIA2	OR2T2	PTK7	SERINC2	TFAP2A	ZNF2	
ASAHI	CCDC129	CYP21A2	FAM83G	HSPG2	KRTAP5-1	MICA	OR2T35	RABL6	SETBP1	TMEM37	ZNF226	
ASB10	CCDC40	CYP2B6	FAM8A1	HTRA1	KRTAP5-8	MKL1	OR2T4	RAD21-AS1	SFRP1	TMEM43	ZNF227	
ASCC2	CCDC66	DACH1	FBF1	ICA1	KRTAP9-1	MKRN1	OR6C76	RAI1	SGK223	TMEM52	ZNF233	

高齢者2名以上でSIFT ”DAMAGING”な変異

アレル頻度1%以下、アミノ酸置換変異でSIFT “DAMAGING” な物として以下の49変異が検出された。

Chromo	Position	Reference	Change	Coverage	Gene_name	Prediction(SIFT)	OMIM Disease
1	6294548	C	G	268	ICMT	DAMAGING *Warning! Low confidence.	
1	22211078	T	G	18	HSPG2	DAMAGING	SCHWARTZ-JAMPEL SYNDROME, TYPE 1
1	145014103	G	C	315	PDE4DIP	DAMAGING *Warning! Low confidence.	
1	145281408	C	T	999	NOTCH2NL	DAMAGING	
1	145281543	C	T	1000	NOTCH2NL	DAMAGING	
1	145281633	C	A	902	NOTCH2NL	DAMAGING	
1	153580072	C	T	308	S100A16	DAMAGING	
2	97877478	G	A	777	ANKRD36	DAMAGING	
2	220037608	A	C	171	CNPPD1	DAMAGING	
2	239037384	G	T	190	ESPNL	DAMAGING	
2	241451294	C	T	88	ANKMY1	DAMAGING	
3	14172424	G	A	451	TMEM43	DAMAGING	ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 5
3	96585671	G	A	85	EPHA6	DAMAGING	
4	159568305	G	T	259	RXFP1	DAMAGING	
5	10649678	C	G	41	ANKRD33B	DAMAGING	
5	10761137	T	G	20	DAP	DAMAGING	
6	15546294	C	T	94	DTNBP1	DAMAGING	HERMANSKY-PUDLAK SYNDROME
6	17602839	C	T	57	FAM8A1	DAMAGING	
7	134931359	G	C	45	STRA8	DAMAGING	
8	100454797	G	T	543	VPS13B	DAMAGING	COHEN SYNDROME
8	143959188	G	T	231	CYP11B1	DAMAGING *Warni	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO STEROID 11-BETA-HYDROXYLASE
9	139735349	C	A	72	RABL6	DAMAGING *Warning! Low confidence.	
9	140777306	C	G	539	CACNA1B	DAMAGING	
10	47087609	G	A	827	PPYR1	DAMAGING	
10	49943983	C	G	277	WDFY4	DAMAGING	
10	75006782	T	G	28	DNAJC9	DAMAGING	
10	75006785	C	G	22	DNAJC9	DAMAGING	
11	2170443	C	T	237	INS-IGF2	DAMAGING *Warning! Low confidence.	
11	66620033	T	C	311	PC	DAMAGING	PYRUVATE CARBOXYLASE DEFICIENCY
11	124253170	G	A	998	OR8B2	DAMAGING	
13	25021263	T	C	352	PARP4	DAMAGING	
13	25021323	A	G	244	PARP4	DAMAGING	
13	26153950	G	T	135	ATP8A2	DAMAGING	
14	94120117	T	G	66	UNC79	DAMAGING	
15	66223280	C	T	547	MEGF11	DAMAGING	
15	75982085	C	T	166	CSPG4	DAMAGING	
16	20442608	C	A	340	ACSM5	DAMAGING	
17	45911840	A	T	393	LRRC46	DAMAGING	
17	48628510	C	A	67	SPATA20	DAMAGING	
17	79254478	A	C	89	SLC38A10	DAMAGING	
19	501714	C	A	70	MADCAM1	DAMAGING	
19	5587268	T	G	67	SAFB2	DAMAGING *Warning! Low confidence.	
19	5587270	A	G	41	SAFB2	DAMAGING *Warning! Low confidence.	
19	36674347	T	C	390	ZNF565	DAMAGING	
19	42905979	G	C	48	LIPE	DAMAGING *Warning! Low confidence.	
22	24579157	G	T	141	SUSD2	DAMAGING	
22	38511670	G	C	41	PLA2G6	DAMAGING	PARKINSON DISEASE 14, AUTOSOMAL RECESSIVE
X	54777555	T	C	178	ITIH6	DAMAGING	
X	140993859	T	A	773	MAGEC1	DAMAGING *Warning! Low confidence.	

結論と展望

- ・ 高齢健常者2名以上で見られる変異は、疾患責任遺伝子の可能性は低い。このためこれらのデータを患者と比較することで、疾患責任遺伝子の同定に役立つことが期待される。
- ・ 検体数を増やして、変異データをコンセンサスを得たうえで一般公開したい（+4検体を計画中）