

平成 23 年度 第一回班会議講演内容

平成22年度厚生労働科学研究費補助金（難治性疾患克服研究事業）

「特発性肺線維症急性増悪及び薬剤性肺障害に関する日本人特異的遺伝素因に関する研究」
(H22 難治 一般 005)

平成23年度第1回班会議

2011.7.22 崎陽軒本店 6F会議室

本研究班の目的

- 特発性肺線維症急性増悪に関する遺伝因子の同定
- 薬剤性肺障害に関する遺伝因子の同定
- 上記の目的を達成するため、患者末梢血DNAを収集、解析する。

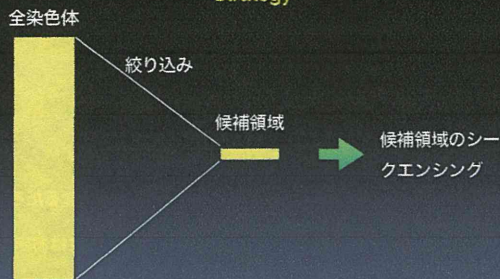
研究スケジュール（工程表）



収集症例

- 2011.07.22現在、データベースに419例登録
- （おそらく）特発性肺線維症急性増悪100例、薬剤性肺障害50例は解析に適した症例
- 250例近くに関しては、臨床情報、画像情報未収集
- 症例数をできるだけ多くしていきたい

Strategy

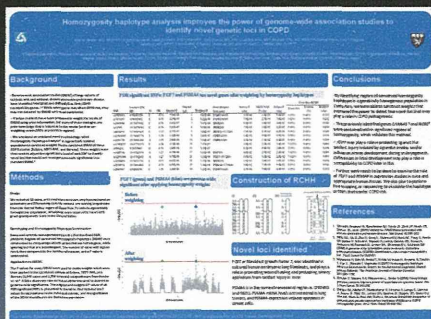


候補領域絞り込みのための手法

- GWAS
 - GWASの限界が明らかになって来ているため、本解析ではあくまで補助的な手法との位置づけ
- Homozygosity haplotype法
- HM on HH法
- Massive parallel sequencingをどのように組み合わせるか

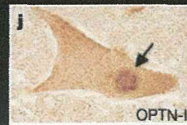
Homozygosity haplotype

Am J Hum Genet 80, 1090–1102, 2007



Thorax in revision

Amyotrophic lateral sclerosis



OTPN gene異常が筋萎縮性脊索硬化症を引き起こす

LETTERS

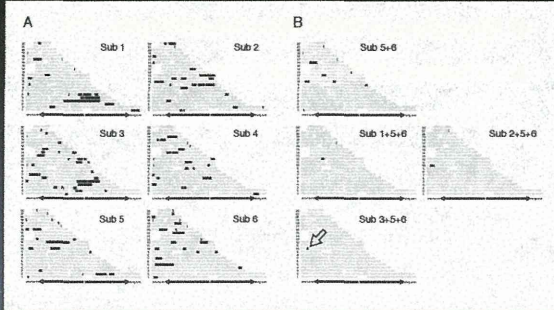
Mutations of optineurin in amyotrophic lateral sclerosis

Hirofumi Matsuyama¹, Hiroaki Morino¹, Hidefumi Ito², Yukihiro Isumi³, Hiromasa Kato⁴, Yasuhiro Watanabe⁵, Yoshimi Kinoshita⁶, Masaki Kamada⁷, Hiroaki Nodera⁸, Hiromasa Suzuki⁹, Osamu Komura¹⁰, Shinya Matsumura¹¹, Kotaro Kobayashi¹², Nobutoshi Morimoto¹³, Kazuhiro Aoki¹⁴, Naoki Suzuki¹⁵, Masashi Kishi¹⁶, Akihito Kanawa¹⁷, Takeshi Hirai¹⁸, Takao Kato¹⁹, Kaumasa Ogasawara²⁰, Asao Hirano²¹, Toshi Takumi²², Hirofumi Kusaka²³, Koichi Higawara²⁴, Ryuji Kaj²⁵ & Hideoshi Kawakami²⁶

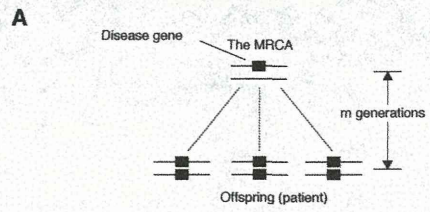
Nature 465:223–226, 2010

qHomozygosity mapping

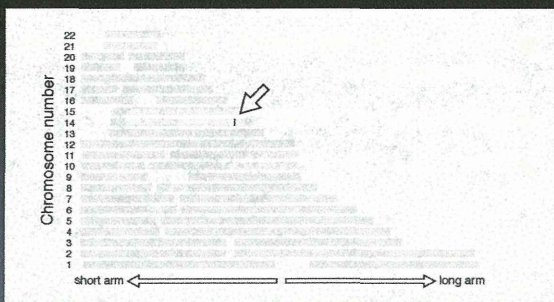
BMC bioinformatics 11 Suppl 7, S5, 2011



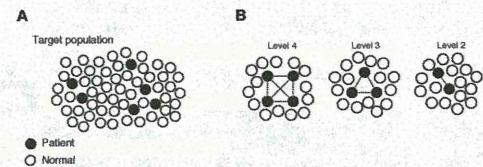
HM on HH analysis



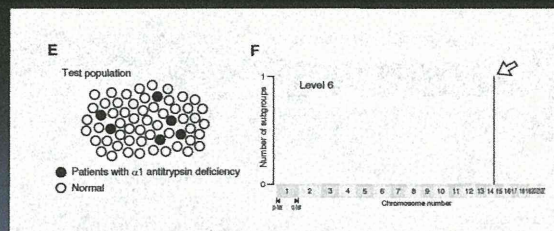
HM on HH analysis



HM on HH analysis



HM on HH analysis



まとめ

- 症例収集の期間はあと一年と思われる
- 順次症例データと画像データの収集をご依頼致します
- ことしの秋から実際のSNPデータでの解析が可能になるだろう
- 収集サンプルを他のゲノム解析手法で解析希望の先生はお申し出ください

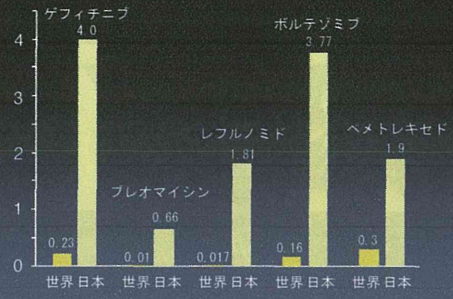
平成23年度 第二回班会議講演内容

特発性肺線維症急性増悪および薬剤性肺障害に関する日本人特異的遺伝素因に関する研究

埼玉医科大学呼吸器内科 萩原弘一

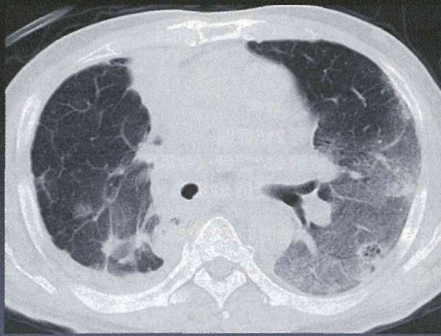
2011.12.17

日本と世界の薬剤性肺障害発生率



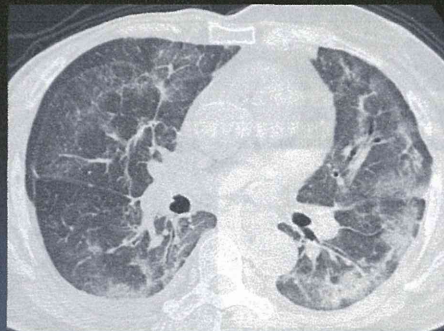
日本人は特定の薬剤で薬剤性肺障害を起こしやすい

症例CT写真



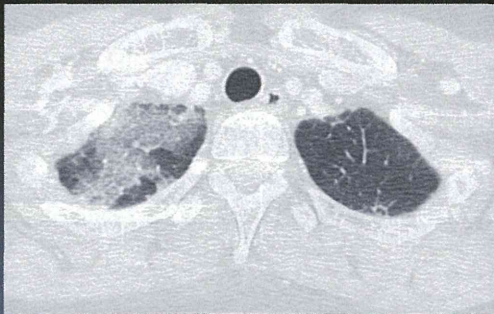
ゲフィチニブ

症例CT写真



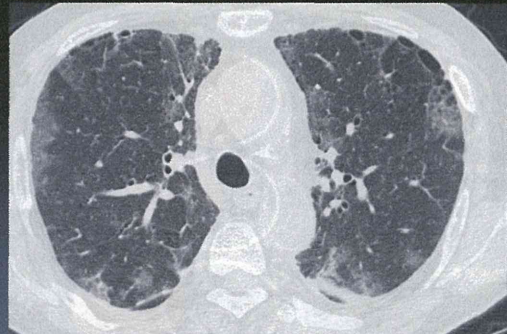
ハクリタキセル

症例CT写真



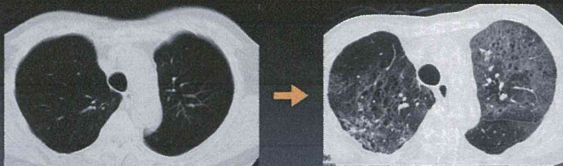
プレオマイシン

症例CT写真



ドセタキセル

症例CT写真

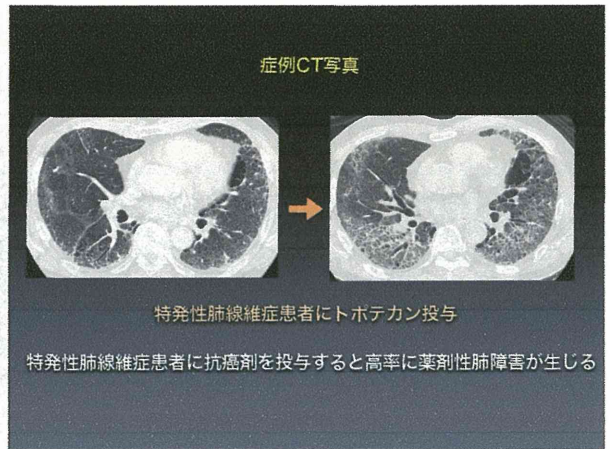
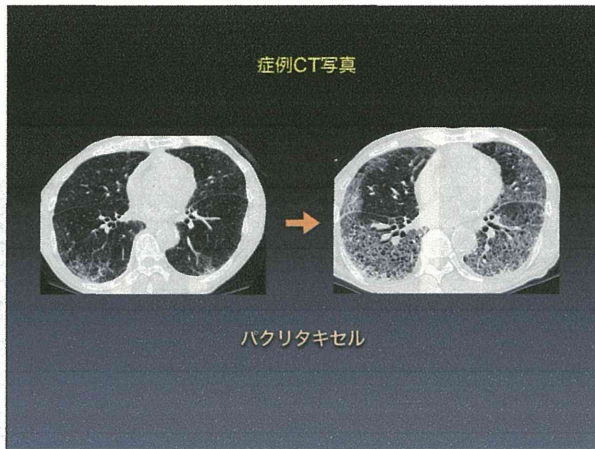


ゲフィチニブ

症例CT写真



UFT



薬剤性肺障害の診断

他疾患を除外したあと、考えられるものが薬剤性肺障害のみの場合に薬剤性肺障害と診断

- 薬剤投与と密接な関係がある発症
- 他疾患（特に間質性肺炎を来す疾患）を除外

SARSによる間質性肺炎（ウイルス性肺炎）

集積症例

おかげさまで、ここまで来ました

- 全集積症例数 433症例
 - 画像収集症例 215症例
 - SNP解析症例 257症例（今年度末）
 - エクソーム解析症例 52症例（今年度末）

画像収集をできるだけ全例：来年度始めまで
「コア症例」に関してSNP解析、エクソーム解析を終了：来年度中頃まで

誤診と集積症例数

臨床診断

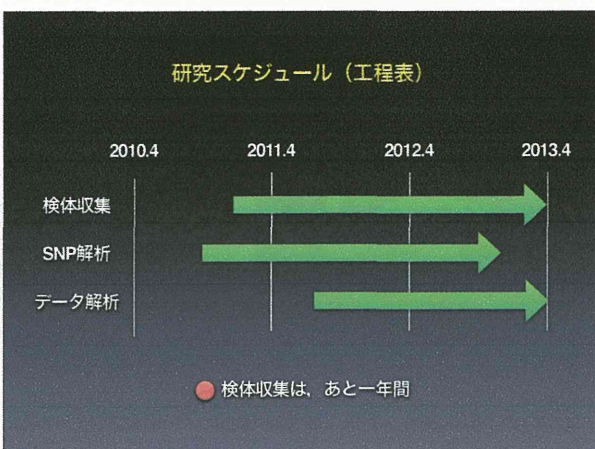
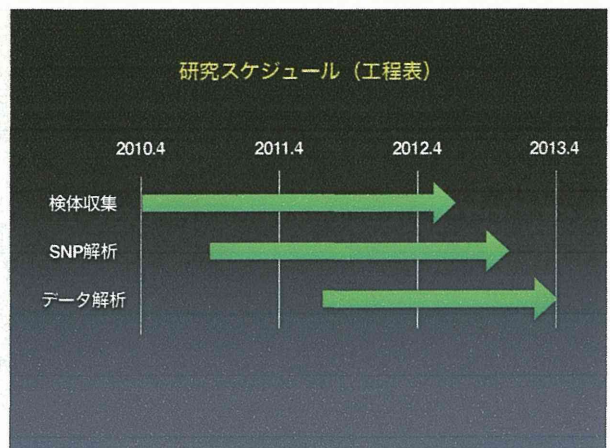
遺伝子Xの多型, 変異

100例程度

300例程度

1000例程度

臨床家を中心とした研究が必須



薬剤性肺障害内訳

● イレッサ+タルセバ	41	● S-1	4
● ドセタキセル	23	● インターフェロン	3
● メトトレキサート	11	● UFT	2
● ハクリタキセル	11		
● アリムタ	8		
● アムルピシン	8		
● アミオダロン	7	● IPF AE	109
● ゲムシタビン	6		
● 漢方薬	6		
● イリノテカン	4		
● レボフロキサシン	4		

SNP症例比較

- 薬剤性肺障害40例 vs control
- 薬剤性肺障害80例 vs control
- 薬剤性肺障害120例 vs control
- IPF急性増悪40例 vs control
- IPF急性増悪80例 vs control
- IPF急性増悪120例 vs control
- 薬剤性肺障害 + IPF急性増悪40例 vs control
- 薬剤性肺障害 + IPF急性増悪80例 vs control
- 薬剤性肺障害 + IPF急性増悪120例 vs control

HH analysis, HHGWAS, HM on HH法で解析します

エクソーム解析

全染色体



絞り込み

候補領域



エクソームシーケンシング
データより参照

エクソームは100例程度収集したい

あと1年間よろしくお願い致します

- 全集積症例数 433症例-> 500症例以上
 - 画像収集症例 215症例-> 500症例以上
 - SNP解析症例 257症例-> 300症例以上
 - エクソーム解析症例 52症例 (年度末予定)
-> 100症例以上

平成24年度 第一回班会議講演内容

特発性肺線維症急性増悪および薬剤性肺障害に関する日本人特異的遺伝素因に関する研究

埼玉医科大学呼吸器内科

萩原弘一、椎原淳、井上慶明、田中知明、宮澤仁志、太田洋充

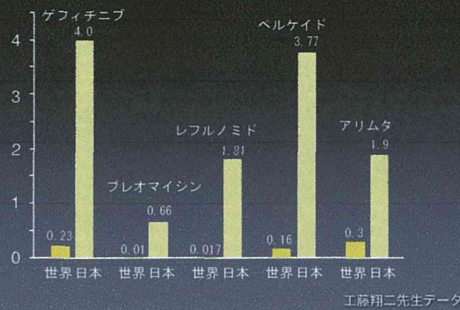
リーカロイヤルホテル（大阪） 2012年7月27日

おわび

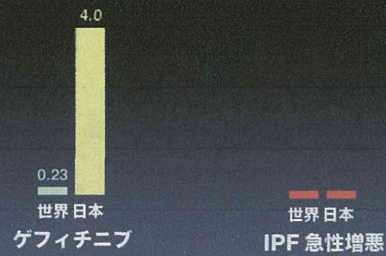
- 萩原研では、最終年度に可能なresourceを全て解析につぎ込みたいと考えています。
- 研究分担者への分担金が0になってしまった先生方が多数になってしまったこと、お詫ひ申し上げます。
- 班会議や、本研究関連の費用は、研究代表者（萩原）から直接お払いいたします。

薬剤肺障害の発生率の比較

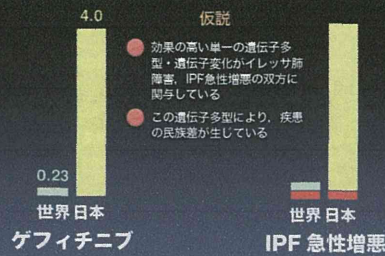
- 日本人は特定の薬剤で欧米人に比べ薬剤性肺障害を著しく起こしやすい



その場合IPFの急性増悪は？

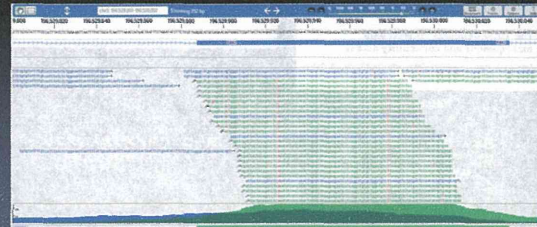


その場合IPFの急性増悪は？

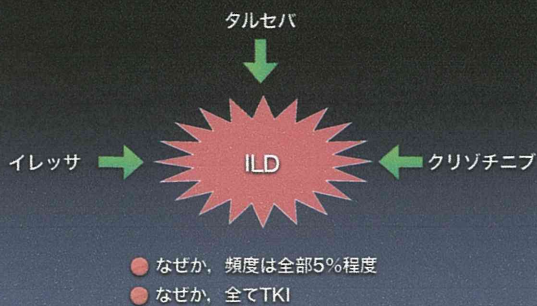


次世代シーケンサーによるエクソーム解析を行っています

- エクソーム解析
 - ・ ヒトの20000遺伝子の配列を全部調べる
 - ・ イントロンなど情報の少ない配列は調べず、エクソンだけ調べる

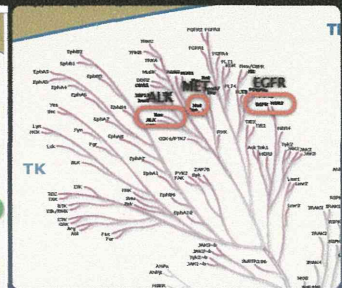
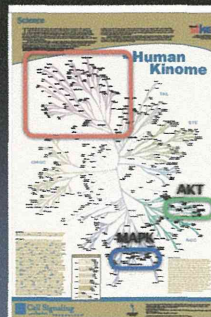


TKIによるILD

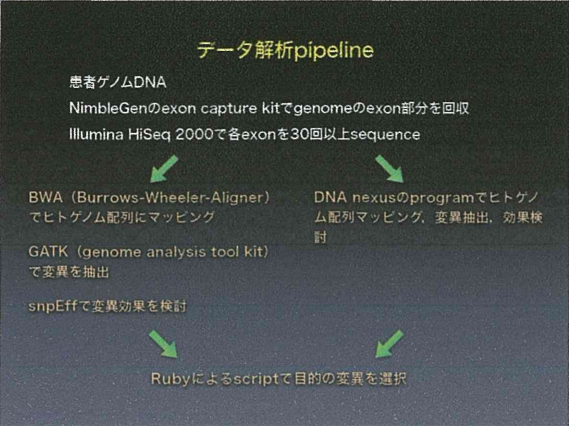
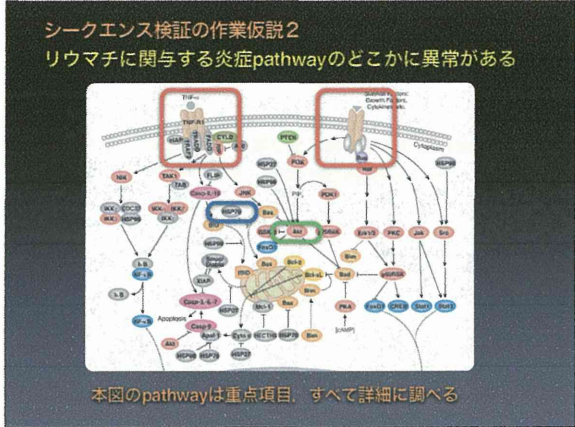
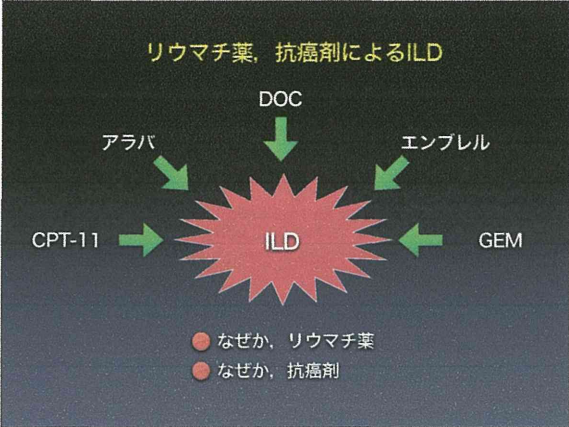


シーケンス検証の作業仮説 1

Kinaseのどれかが日本人で機能不全を起しているのでは？



Kinaseは重点項目、すべて詳細に検討



一人当たりアミノ酸置換が数千個〜一万数千個

chr1	69511	OR4F5	T	A
chr1	865694	SAMD11	H	Y
chr1	888859	NOC2L	I	V
chr1	909238	PLEKHN1	R	P
chr1	966933	AGR1	H	P
chr1	1115555	TLL10	P	H
chr1	1115805	TLL10	A	T
chr1	1222267	SCNN1D	R	P
chr1	1392652	ATAD3C	R	G
chr1	1421991	ATAD3B	R	Q
chr1	1423263	ATAD3B	K	R
chr1	1423266	ATAD3B	R	G
chr1	1430985	ATAD3B	R	C
chr1	1431165	ATAD3B	P	S
chr1	1455652	ATAD3A	A	S
chr1	1551926	MB2	SF	SL
chr1	1558792	MB2	M	T
chr1	1599812	SLC35E2	V	I
chr1	1647901	CDK11A		KE

民族差検証のための対照データは？

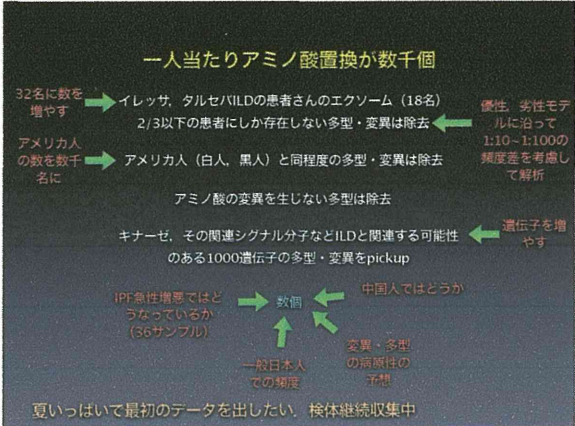
1000 Genomes project

Nature Methods 9:1-4, 2012

日本人（東京）、中国人（北京）、コーカシアン、アフリカ系アメリカ人など各民族数十名ずつ

NHLBI Exome Sequencing Project

ヨーロッパ系アメリカ人、アフリカ系アメリカ人合計6500名



平成 24 年度 第二回班会議講演内容

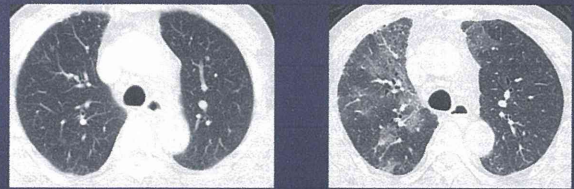
平成24年度 第二回班会議講演内容

特発性肺線維症、薬剤性肺障害に
関与する日本人特異的遺伝素因に
関する研究
2012年度第2回班会議

2012.11.8 ダイワロイネットホテル岡山駅

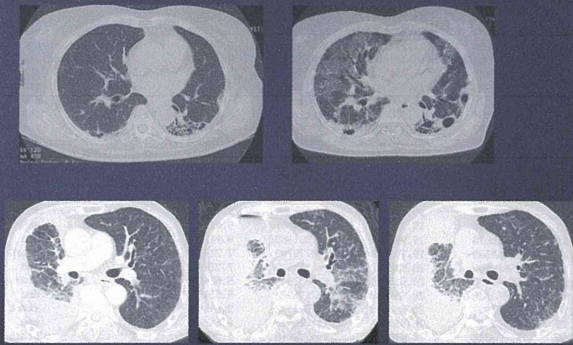
椎原 淳、萩原 弘一

薬剤性肺障害

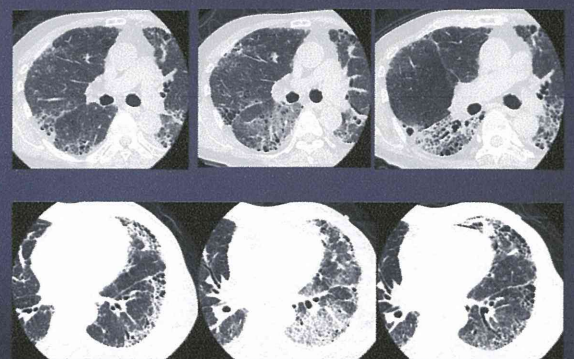


Gefitinib使用2ヶ月

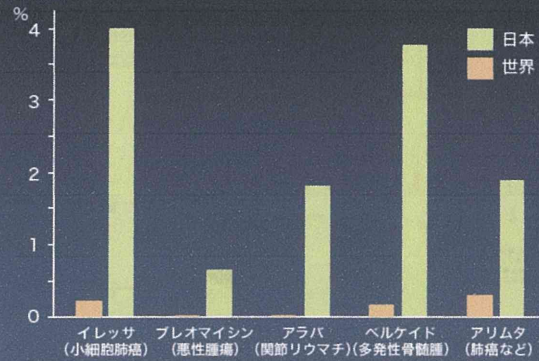
Gefitinib / Erlotinib



IPF Acute Exacerbation



薬剤投与者
中の頻度



日本人肺の脆弱性は日本人特有の遺伝因子によるものではないか？

- 薬剤性肺障害だけでなく、
- ・ 特発性肺線維症の急性増悪
- ・ 間質性肺炎合併患者の、術後増悪
- ・ Amyopathic DMにおける、急速進行性の間質性肺炎

も日本人に極めて多い。



厚生労働省「難治性疾患克服研究事業」
「特発性肺線維症急性増悪及び薬剤性肺障害に関与する日本人特異的遺
伝素因に関する研究」(平成22~24年度) 研究代表者 萩原弘一

Interstitial lung disease collected : 549

Cases that clinical images are available : 279

Core cases : 183

Drug induced ILD : 118

Gefitinib / Erlotinib / Crizotinib : 38

DOC : 22

MTX : 11

PTX : 7

Amiodarone : 5

漢方 : 5

PEM : 5

CPT : 4

GEM : 3

IPF Acute Exacerbation : 65

Gefitinib / Erlotinib / Crizotinib : 38

SNP analysed : 32

Exome analysed : 38

DOC : 22

SNP analysed : 19

Exome analysed : 15

IPF Acute Exacerbation : 65

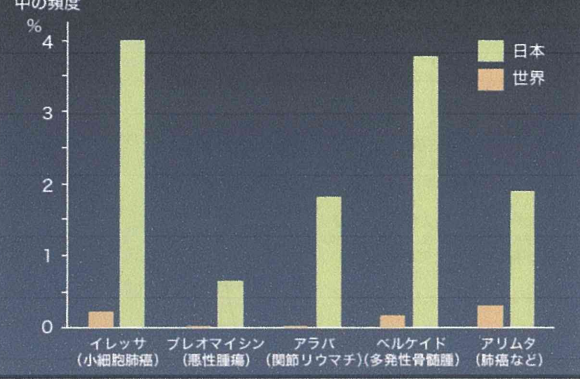
SNP analysed : 62

Exome analysed : 45

サンプル毎にSNP, InDelといったmutationをtxt形式で打ち出し。
Coding sequence, non-synonymousを中心に拾い上げて行く

Table with columns: Chr, Position, Ref, Alt, Info, ESP, EA, AA. It lists various genetic variants across different chromosomes and populations.

薬剤投与者
中の頻度



事実

- 日本人における、イレッサによるILD発症頻度は5%前後
欧米人ではほぼ皆無
中国人でも症例の報告は極めて乏しい

全てのSNP, InDelに関して、上記に合致するものをリストアップしていく

それぞれのSNPの保有頻度を、欧米、中国、日本人、疾患群の頻度と比較して絞り込んで行くstrategy

Table with columns: Chromosome, Position, Ref, Alt, Info, ESP, EA, AA. It provides detailed variant information and population frequencies.

Summary table showing variant frequencies across different population groups: CEU, JPT, CHB, IreTar, IFF.

比較対象その1 : ESP (Exome Sequencing Project) から 6500人分のEA (European American) と AA (African American) の SNPサマ리를ダウンロードして比較

NHLBI Exome Sequencing Project (ESP) Exome Variant Server. Includes navigation links like Home, Data Browser, and a list of downloadable files.

比較対象その2 :

日本人、中国人のデータベースは1000人ゲノムプロジェクト (http://www.1000genomes.org) からダウンロードし、比較

1000 Genomes A Deep Catalog of Human Genetic Variation. Includes variant frequency data for different populations and a note about excluding certain groups.

絞り込み方法

- Pa: 健康日本人の遺伝子異常保有頻度がa%以上である確率
Pb: 疾患群の遺伝子異常保有頻度がb%以上である確率
Pc: 健康日本人の遺伝子異常保有頻度が、欧米人 (EA, AA) のc倍以上である確率
Pd: 健康日本人の遺伝子異常保有頻度が、中国人 (CHB) のd倍以上である確率

P = Pa x Pb x Pc x Pd

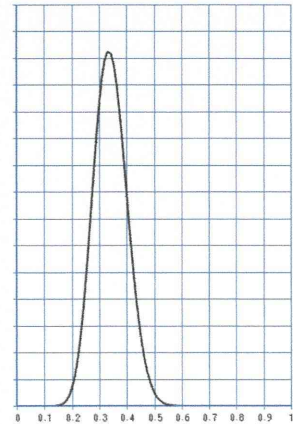
として、このPの値が大きい方から順に検討していく

一般人口n人を調べ、r人が遺伝子Aを有することが分かった場合、一般人口中の遺伝子Aの頻度xを表す分布の確率密度関数f(x)は

f(x) = Be_x(1+r, 1+n-r) = x^r(1-x)^{n-r} / B(1+r, 1+n-r) = x^r(1-x)^{n-r} / integral_0^1 t^r(1-t)^{n-r} dt

というベータ分布で表される。

右に、一般人口60人を調べて、20人が遺伝子変異Aを有する場合のベータ分布の曲線を例示する。



このとき、遺伝子Aの頻度の分布の累積密度関数は不完全ベータ関数となる

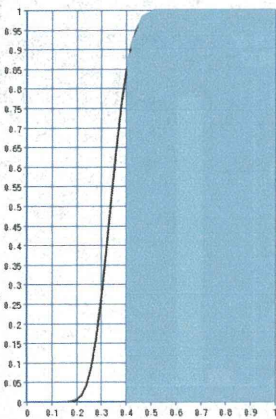
$$I_x(1+r, 1+n-r) = \frac{\int_0^x t^r (1-t)^{n-r} dt}{\int_0^1 t^r (1-t)^{n-r} dt}$$

右に、一般人口60人を調べて、20人が遺伝子変異Aを有する場合は不完全ベータ関数を例示する。

上記の「一般人口でn人を調べ、r人が遺伝子Aを有する」場合、一般人口中の頻度が例えば0.4以上である確率Pは

$$P = 1 - I_{0.4}(1+r, 1+n-r)$$

で得られる。



比較対象その1: ESP (Exome Sequencing Project) から 6500人分のEA (European American) と AA (African American)の SNPサマリをダウンロードして比較

NHLBI Exome Sequencing Project (ESP)
Exome Variant Server

Home Data Browser Data Usage and Release How to Use What's New Contact and FAQ Downloads

The following downloadable files contain SNPs and coverage data for the ESP 6500 exomes (chromosomes 1-22, and X).

ESP6500.snps.txt.tar.gz (SNPs and annotations in space-delimited text format)
 ESP6500.snps.vcf.tar.gz (SNPs and annotations in VCF format - same data as ESP6500.snps.txt.gz)
 ESP6500.coverage.all_sites.txt.tar.gz (sequencing coverage data for every site in our targets)
 ESP6500.coverage.summary.blocks.txt.tar.gz (sequencing coverage summary data for every sequencing block in our targets)

Chromosome	Position	Ref	Alt	Info	WASHP	ESP EA	ESP AA
1	14610	T	C	(repeats)SNP rs79134172	4300	0	2203
1	14653	C	T	(repeats)SNP rs82636297	4300	0	2203
1	14677	G	A	(repeats)SNP rs111299168	4300	0	2203
1	14696	A	C	(repeats)SNP	4300	0	2203
1	14783	TC		(repeats)DEL	4300	0	2203
1	14786	C	T	(repeats)SNP	4300	0	2203
1	14795	GG	CA	(repeats)SNP	4300	0	2203
1	14975	C	T	(repeats)SNP rs1120877;rs11883341	4300	0	2203
1	14976	G	A	(repeats)SNP rs71252251;rs77117775	4300	0	2203
1	15205	G	A	(repeats)SNP	4300	0	2203
1	16957	G	T	(repeats)SNP	4300	0	2203
1	69322	T	A	SNV	4300	0	56
1	69428	T	G	SNV rs1149759101	GRF5 3203	123 92	1898 12

P_a: 健康日本人の遺伝子異常保有頻度がa%以上である確率
 P_b: 疾患群の遺伝子異常保有頻度が、b%以上である確率
 P_c: 健康日本人の遺伝子異常保有頻度が、欧米人(EA, AA)の4倍以上である確率
 P_d: 健康日本人の遺伝子異常保有頻度が、中国人(CHB)の4倍以上である確率

疾患群サンプル、健康日本人サンプル(1000人ゲノム)、健康中国人サンプル(1000人ゲノム)は、サンプル数が少ないので既述のベータ分布などを用いて、頻度を確率学的に算出

欧米人(EA, AA)はサンプル数が多いので、絶対値で算出する。

$$P = P_a \times P_b \times P_c \times P_d$$

を尺度として、このPの値が大きいく方から順に検討していく

遺伝形式は、隐性タイプが有力であり、hg19のreferenceではないalternativeのSNP、Indelが有力である。(reference側が疾患原因SNPということもあり得る)

疾患原因遺伝子の発症形式が優性で、かつAlternativeが原因とした場合の、尤度尤度

上位30程度で、全体の95%を占める
 ...この中に95%の確率で原因遺伝子変異が存在する

P	sumP	proportion	sum proportion	chr	reference	alternative	type	rs#	gene
0.499970352	4.714289148	0.105415641	0.105415641	22	A	AC	INDEL	rs74910017	IRBC2
0.467101389		0.10322264	0.208638281	4	TTTTG	T	INDEL	rs14001938	DCHS2
0.469316679		0.099549839	0.30818812	12	T	G	SNV		FAM188A
0.389424864		0.08207701	0.39026513	11	TTCGTCCTG	T	INDEL	rs141871766	MAM2
0.342782605		0.07121217	0.46307688	11	C	GTGGCCTAC	INDEL	rs11208490	DNAH1
0.179077241		0.039181407	0.50122827	12	T	G	SNV	rs12297853	FAM188A
0.178200242		0.037374899	0.538613276	7	AT	A	INDEL	rs146697385	WDR6
0.173640618		0.036832083	0.575445359	6	CCTG	C	INDEL	rs5981170	MAMP3K4
0.171714706		0.036422032	0.611869391	12	A	G	SNV	rs10878022	FAM188A
0.140170358		0.029718022	0.641587413	7	TD		INDEL	rs34061445	KCP
0.139228869		0.02933996	0.671127884	12	T	A	SNV		FAM188A
0.13917386		0.02921108	0.700648964	18	A	ATCT	INDEL	rs34990489	GREB1L
0.138089308		0.028292513	0.728941477	11	CA	C	INDEL		MUC2
0.132340942		0.028071705	0.758013182	19	GGGTCTCTG	G	INDEL	rs67168176	FAM71E2
0.128976442		0.027370156	0.785383338	11	TAG	T	INDEL	rs115897236	RP11-197M16.4
0.123181477		0.026124589	0.811707827	18	G	C	SNV	rs7199961	ZNF469
0.110918341		0.023927617	0.835635444	12	C	A	SNV	rs11114468	PTRFQ
0.090259191		0.018787487	0.854422931	3	T	C	SNV	rs9534243	C5orf77
0.090152559		0.017978364	0.872401295	12	A	A	SNV		OXYTL1
0.059194087		0.012559046	0.732384311	5	TGCA	T	INDEL	rs110563854	ANKRD31
0.049082608		0.01020053	0.883584861	1	G	A	SNV	rs53535390	MUC12
0.040175338		0.008521854	0.892106715	12	C	GGGGCCTGC	INDEL	rs15119723	NCOR2
0.028948425		0.006313577	0.898420292	7	T	C	SNV	rs10909022	WDR6
0.027498719		0.005104842	0.903525134	8	G	A	SNV		C5orf77
0.024104852		0.002348452	0.910873586	12	G	A	SNV	rs34782871	FAM188A
0.021913478		0.001798441	0.912671108	8	T	A	SNV	rs13235053	RP11
0.020077489		0.00170003	0.923371138	12	T	G	SNV		FAM188A
0.020398806		0.001387719	0.924758957	21	GA	G	INDEL	rs25393002	AP020244.1
0.024023062		0.002067447	0.933826405	12	TGG	T	INDEL	rs5839209	ZNF384
0.02442221		0.001760255	0.93558666	7	G	C	SNV	rs35257244	MUC12
0.021625919		0.004813451	0.94039811	21	A	G	SNV	rs982735	AP020244.1

今後の課題

- mapping (得られたバラバラのPCR productをreference genomeに当てはめていく作業) におけるエラーの除去
- 次世代シーケンサー、およびデータ解析上のエラーの除去
- いかに「最後の1つ」に落とし込んで行くか
- PCRを実際に全てのサンプルに行い、SNP/Indelの存在を確認する

遺伝学的解析プログラム集

各民族での多型頻度のベイズ理論による計算

一般人口で n 人を調べ、 r 人が遺伝子 A を有することが分かったとき、一般人口中の遺伝子 A 頻度を表す分布の確率密度関数 $f(x)$ は、

$$f(x) = Be_x(1+r, 1+n-r)$$

になる。ここで $Be_x(\alpha, \beta)$ は（第一種）ベータ分布である。Excel には関数があり、
 $Be_x(1+r, 1+n-r) = \text{BETA.DIST}(x, 1+r, 1+n-r, \text{FALSE})$

なお、

$$Be_x(\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} B(\alpha, \beta)$$

分母の $B(\alpha, \beta)$ はベータ関数であり、

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

遺伝子 A の頻度の分布の累積密度関数は $I_x(\alpha, \beta)$ 正則化された不完全ベータ関数となる

$$I_x(\alpha, \beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt}$$

$I_x(\alpha, \beta)$ も excel に組み込み関数がある。

$$I_x(\alpha, \beta) = \text{BETA.DIST}(x, \alpha, \beta, \text{TRUE})$$

上記の「一般人口で n 人を調べ、 r 人が遺伝子 A を有する」場合、一般人口中の頻度が 0.05 以上である確率 P は、

$$P = 1 - \text{BETA.DIST}(0.05, 1+r, 1+n-r, \text{TRUE})$$

となる。

優性遺伝モデルでは Homo+Hetero を遺伝子 A 保有者、劣性遺伝子モデルでは Homo のみを遺伝子 A 保有者と読み替える。これを、reference, alternative に関して行う。

ESP 集団はサンプル数が大きいので、ESP 集団との比較は上記の式を直接使用する。例えば、European American での遺伝子 A の頻度が a であり、日本人 n 人を調べ、 r 人が遺伝子 A を有することが分かったとき、日本人での遺伝子頻度が European American の 10 倍である確率 P は、

$$P = 1 - \text{BETA.DIST}(10a, 1+r, 1+n-r, \text{TRUE})$$

で計算できる。

サンプル数の少ないもの同士、例えば CHB と JPT での比較が必要な場合は、以下のようなやり方が必要かもしれない。

CHB での遺伝子 A の頻度が A であり、CHB N 人を調べ、 R 人が遺伝子 A を持っていた。JPT での遺伝子 A の頻度が a であり、JPT n 人を調べ、 r 人が遺伝子 A を持っていた。ここで、日本人での遺伝子 A 頻度が CHB での遺伝子 A 頻度の 10 倍以上の確率は、 $0 \leq x \leq 1$ の区間を Δx に細分化して（例えば 0.01 の 100 区間に分けて）考える。

$(x, x + \Delta x)$ の区間の CHB の確率は

$$Be_x(1+R, 1+N-R)\Delta x$$

または、

$$I_{x+\Delta x}(1+R, 1+N-R) - I_x(1+R, 1+N-R)$$

と書ける。この場合 JPT の確率が 10 倍以上となる確率は、

$$1 - I_{10x}(1+r, 1+n-r)$$

よって、 $(x, x + \Delta x)$ の区間で JPT が CHB の 10 倍以上となる確率は、

$$(I_{x+\Delta x}(1+R, 1+N-R) - I_x(1+R, 1+N-R))((1 - I_{10x}(1+r, 1+n-r)))$$

これを、 $0 \leq x \leq 1$ （実際は x は 0.1 以上を取れないので $0 \leq x \leq 0.1$ ）の範囲で足し合わせる。積分が難しいので、実際に加えてしまう方が簡単であろう。

