

がんの個別化医療とスーパーコンピュータ

宮野 悟

東京大学医科学研究所ヒトゲノム解析センター

miyano@ims.u-Tokyo.ac.jp

新生命科学分野開拓とスーパーコンピュータ「京」

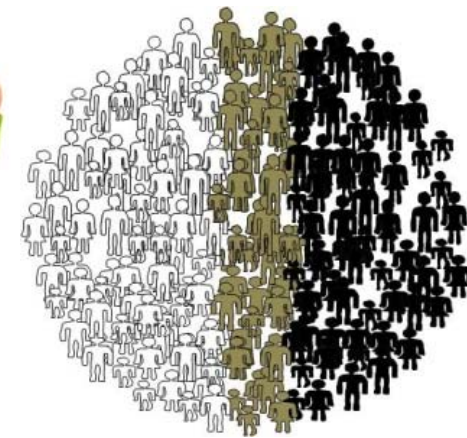
九州大学医学部百年講堂 中ホール1・2

2013年9月19日

「世界に一つだけの花」
私たちは一人一人異なる
ゲノムを持っている



「がんは日本の国民病」日本人
の半分が罹り、3分の1が亡く
なっている



歳をとらない人 その後ろで歳をとっていった人



磯野カツオ(11歳、小学5年生)
(テレビ番組「サザエさん」より。1969年10月放送開始)



声の出演
高橋和枝
(2代目:28年間)
1970年1月11日~1998年5月10日

つまり、
『年寄りはずらいよ』

骨髄異形成症候群で、1999年3月23日
東大医学部附属病院で亡くなる(70歳)。

➤ どんな病気？

- 骨髄で正常な血液を作れなくなる、代表的な「血液がん」。
- 急性骨髄性白血病への移行が特徴。
- お年寄りに多い。日本で、推定で毎年新たに5千人が罹患。
- **その原因は不明だった。**

➤ 治療法は？

- 骨髄移植しかない。
- しかし、骨髄移植の適応は一般に60歳まで。

みんな歳をとる

60歳、65歳を過ぎても、なんとか働き続け、互いに支えていかねばならない日本社会が到来する。そう簡単にがんで寝付くわけにはいかない！

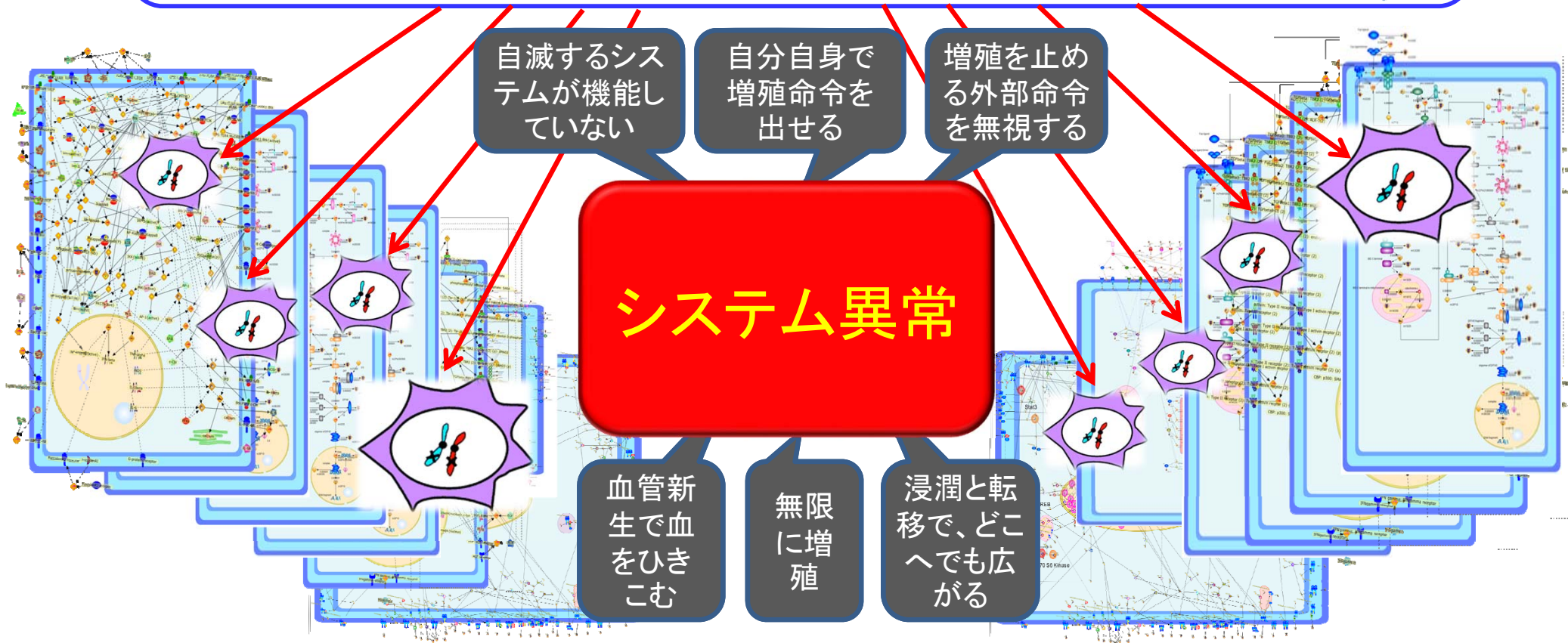
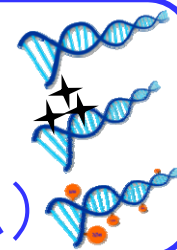


読売新聞 2012年1月30日(月)より

私はなぜがんになるの？



- 個人個人で異なっている遺伝的要因(DNA)
- 腫瘍細胞に蓄積した遺伝子変異(がんゲノム)
- 環境要因などによるゲノム修飾の変異(エピゲノム)



がんの悪性度や治療応答性、副作用の出やすさなどを規定している

さらに、がんは複雑

- 原発から浸潤、転移へ、がん細胞は変幻。
- がん幹細胞や、正常細胞である血管内皮細胞、免疫炎症細胞などと「つるみ」ながら、
- 薬剤耐性を獲得し、
- 時空間で進化するヘテロな細胞集団
- 複数の原発が進化することもある！

さらに、がんは複雑



気持ちは、ムンクの『叫び』
この複雑さをどのように理解すればよいのか？

スパコンで加速できた がん研究の事例

Breaking the Complexity of Cancer

数学とスパコン
を駆使した
大規模データ解析と
数理モデリング

システム生物学・バイオ
インフォマティクス・
遺伝統計学

融合

腫瘍学・がん病理学・が
ん生物学

大規模網羅的
オミックス解析

がんの多様性
とダイナミズム
をデジタル化

システムがん



システムでがんを読み解く



新学術領域研究 システムがん (複合領域: 4201)

システムの統合理解に基づくがんの先端的診断、治療、予防法の開発

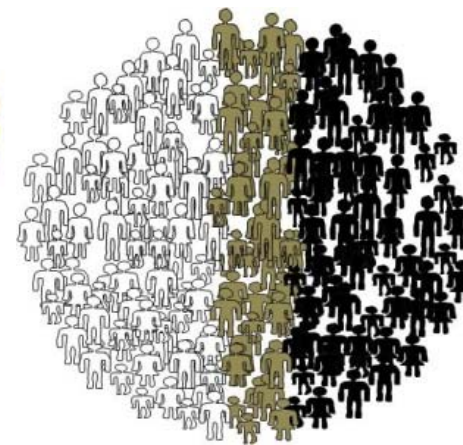
領域代表 宮野 悟 PhD
A01班長 稲澤譲治 MD PhD
A02班長 高橋 隆 MD PhD

東大医科学研究所ヒトゲノム解析センター
東京医科歯科大難治疾患研究所
名古屋大学医学研究科

「世界に一つだけの花」
私たちは一人一人異なる
ゲノムを持っている



「がんは日本の国民病」日
本人の半分以上が罹り、3分の
1が亡くなっている



融合研究が活発化状態

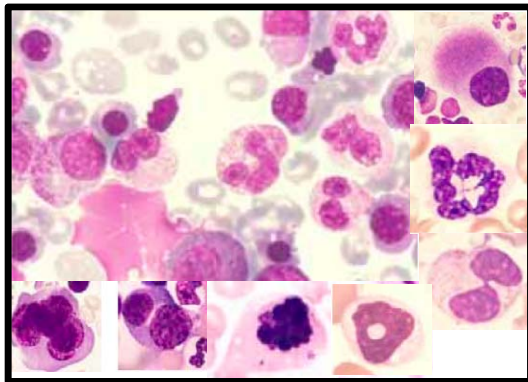
システムがん研究の方法論が導入され、本領域を象徴する成果が出てきた。

がん研究の歴史に刻まれる発見
(Nature 2011)

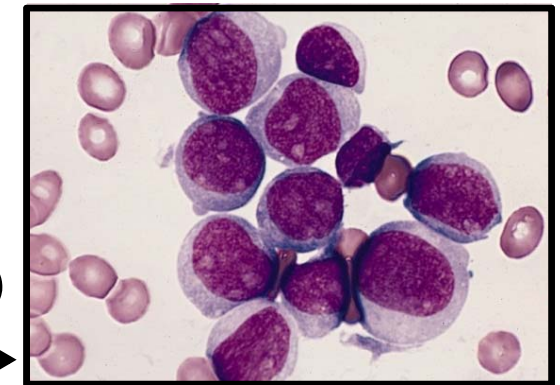
「骨髄異形成症候群(MDS)の解明」
大規模次世代シーケンサー解析、スパコンをフル活用した数理解析チームとの共同により、研究開始から論文発表までわずか1年ほどで(2010年7月～2011年9月11日)、未解決の大問題を解決・発表。

骨髄異形成症候群 (MDS) の 原因遺伝子の発見

難治性血液がんの克服に向けて



骨髄異形成症候群



急性骨髄性白血病

A02 小川誠司
(東大医学部附属病院)
(現在、京都大学医学研究科教授)

移行

◇いろいろな血液細胞が
作れなくなることで
赤血球減少 → 貧血、動悸、息切れ

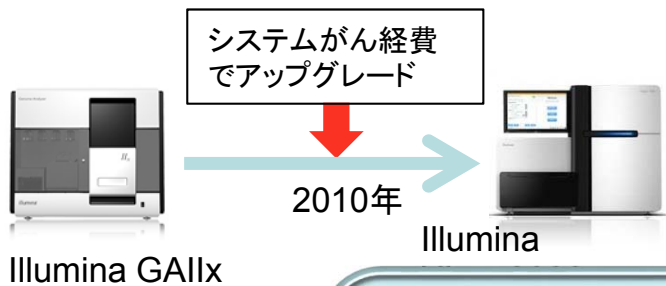
白血球減少 → 感染しやすくなる

血小板減少 → 出血しやすい

MDS: myelodysplastic syndromes

【共同研究】

A01 宮野 悟 (東大医科学研究所)



スパコンと次世代シーケンサーを使い、がんの新たな原因機構を発見した我が国初の大規模ゲノムシーケンス研究

臨床的・生物学的検証

MDSの原因は不明

次世代シーケンサーHiSeq2000を用いて、29検体の全エクソン・シーケンスを行う

候補変異を高速リシーケンス技術を用いて、多数検体について効率よくスクリーニングする技術を開発

- **582例の検体**を用いた変異解析で再現確認。
- 変異体を導入HeLa細胞でRNAスプライシング異常が起こることを確認。
- 変異したスプライシング因子が、造血機能に及ぼす効果をマウスに骨髄移植して検討。

ヒゲノム解析センタースパコン上に、超高速、効率的に変異遺伝子の候補を検出する耐故障性データ解析パイプラインを構築

高精度変異同定アルゴリズムを開発。血液腫瘍試料には正常細胞が混入しているという困難を克服。

268個の体細胞変異を同定

RNAスプライシング経路の4つの遺伝子に異常が頻出

がん研究の歴史で初めての知見

原因を解明

2011年9月

データ解析パイプラインを公開



6000コア (75TFLOPS)
1PB高速ストレージ

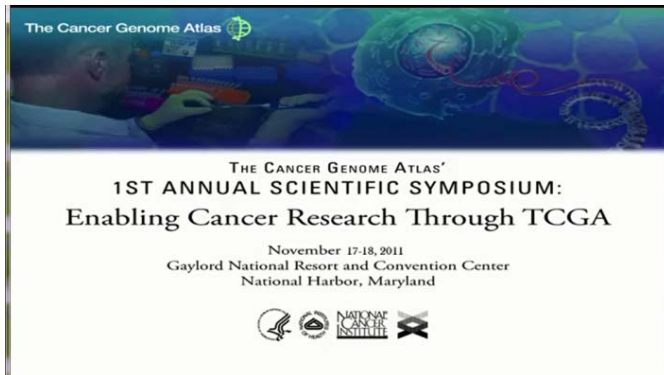


エクソーム解析・変異解析

融合遺伝子解析



がん研究の歴史に刻まれる発見



TCGA: 2005年に開始。これまでに毎年約50M\$の研究費



Keynote Speechで紹介(2011年11月17日)

Eric Lander, Director of Broad Institute of MIT and Harvard
 Short Biography
 BS: Mathematics
 MS: MBA
 PhD: Molecular Biology

Insights from Genomic Approaches	
Functional classes	
Protein and lipid kinases	BRAF (melanoma), PIK3CA (breast), PIK3R1 (GBM), EGFR (lung), FGFR2 (endometrial), JAK2 (myeloproliferative)
Lineage survival genes	MITF (melanoma), NKX2-1 (lung), SOX9 (colon), SOX2 (squamous lung and esophageal),
Epigenetic regulators	DNMT3A (AML), EZH2 (DLBCL), UTX (many), MLL2/3
Metabolic enzymes	IDH1/2 (GBM and others)
RNA splicing factors	SRSF2, SF3B1, ZRSR2, U2AF1 (MDS), SF3B1 (sideroblastic anemia, CLL)
Translocations	ERG (prostate), ALK (lung)
Surprises	
Notch	Oncogene (T-ALL) and tumor suppressor (squamous)



NATURE | ARTICLE [previous article](#) [next article](#)

Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida, Masashi Sanada, Yuichi Shiraiishi, Daniel Nowak, Yasunobu Nagata, Ryo Yamamoto, Yusuke Sato, Aiko Sato-Otsubo, Ayana Kon, Masao Nagasaki, George Chalkidis, Yutaka Suzuki, Masashi Shiosaka, Ryoichiro Kawahata, Tomoyuki Yamaguchi, Makoto Otsu, Naoshi Obara, Mamiko Sakata-Yanagimoto, Ken Ishiyama, Hiraku Mori, Florian Nolte, Wolf-Karsten Hofmann, Shuichi Miyawaki, Sumio Sugano, Claudia Haeflrich, H. Phillip Koeffler, Lee-Yung Shih, Torsten Haeflrich, Shigeru Chiba, Hiromitsu Nakauchi, Satoru Miyano & Seishi Ogawa

[Show fewer authors](#)

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Nature 478, 64–69 (06 October 2011) | doi:10.1038/nature10496
 Received 07 June 2011 | Accepted 24 August 2011 | Published online 11 September 2011

この発見は、世界で初めて、骨髄異形成症候群(MDS)の原因遺伝子を発見したという意義だけでなく、「RNAスプライシング」の異常が、がんの発症に関わることを示した世界で初めてのもの

My DNAを「シーケンス」

生物のDNA情報を読み取る装置は一般にシーケンサーとよばれ、A, T, C, Gの文字で綴られるDNA情報(ヒトの場合30億文字の情報)をコンピュータで読めるように取り出すことを「シーケンス」とよんでいます。

だれもが自分のDNA情報を利用できる時代が到来

現在2013年

27時間、40万円
装置：1億円以上
リード長：100-150



Illumina HiSeq2500

限界 シークエンス量
に線形比例して高価
な試薬が必要

2013年～2014年

6時間、10万円
装置：1500万円
リード長：400b



Ion Proton

2014年

1時間、10万円～
装置：300-6000万円
リード長：50Kb-100Kb



Oxford Nanopore Technologies
2000ナノポア集積

2014年～

1時間以内、1万円
装置：10万円
リード長：1Mb



Genia Technologies
100万ナノポア集積
COMSチップ→数百円

2015年～

超安価・高速・
高精度

IBMなどを初め
とする様々な技術
開発と応用への
展開
ゲノム及びそれ
に由来する情報の
臨床翻訳・解釈
が最も重要な
課題

★ シリコンシーケンサーの登場
シリコンチップのコストはとても安い

さらに、トランスクリプトーム、
エピゲノムもこの装置で解析

しかし、DNAビッグデータを解読して
活用できる人材・計算資源が不足

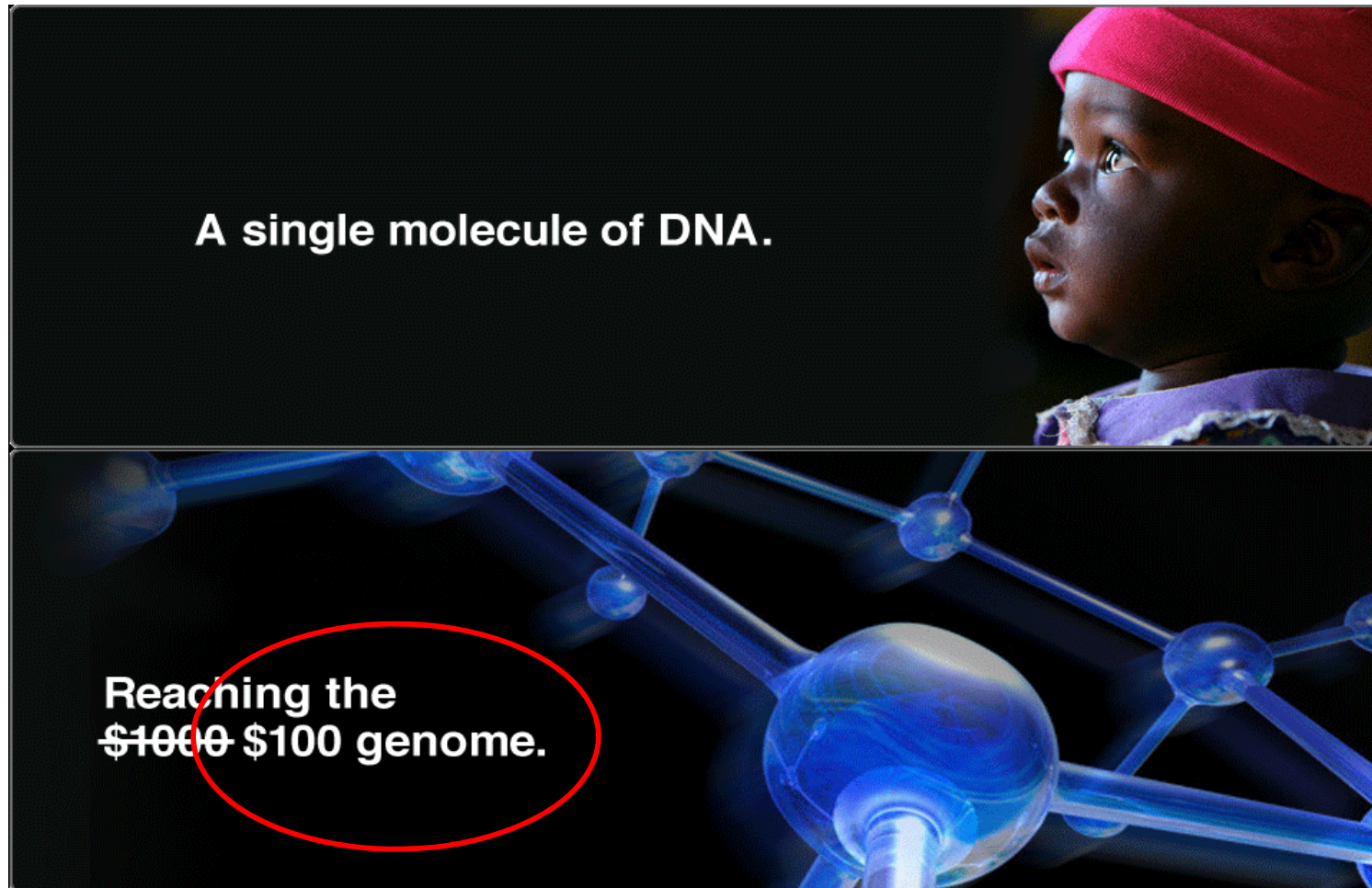
パーソナルゲノム時代のはじまり

ゲノムビッグデータの誕生



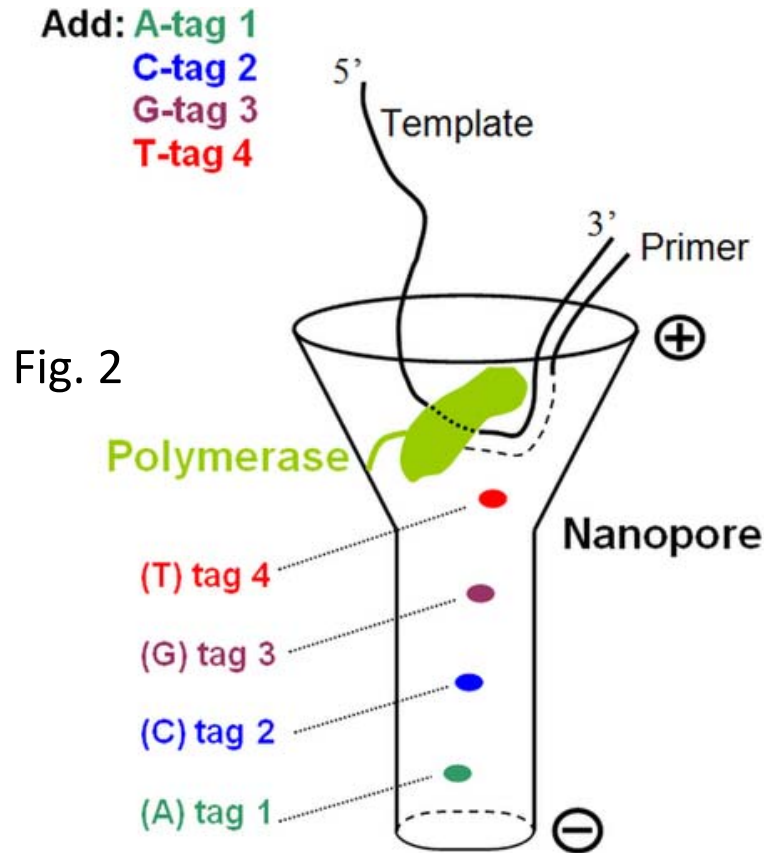
Genia Technologies, Inc.

2013年～ヒトゲノムは100ドル以下、1時間以内
“Last-Generation Sequencer”



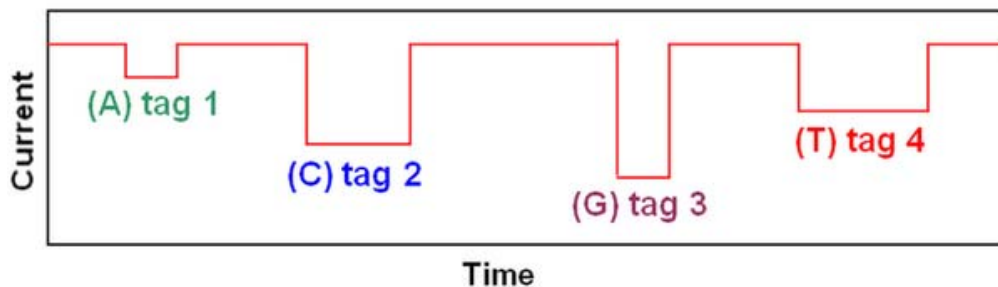
<http://www.geniachip.com/>

DNA sequencing by a nanopore with phosphate-tagged nucleotides



General principle of single molecule electronic DNA sequencing by synthesis using PEG-labeled nucleotides and nanopore detection

The single molecule electronic Nano-SBS system, which is shown schematically in [Fig. 2](#), depicts the DNA polymerase bound in close proximity to the nanopore entrance. A template to be sequenced is added along with the primer. To this template-primer complex, four differently tagged nucleotides are added to the bulk aqueous phase. After polymerase catalyzed incorporation of the correct nucleotide, the tag-attached polyphosphate will be released and pass through the nanopore to generate a unique ionic current blockade signal, thereby identifying the added base electronically because the tags have distinct chemical structures.



Kumar S, Tao C, Chien M, Hellner B, Balijepalli A, Robertson JW, Li Z, Russo JJ, Reiner JE, Kasianowicz JJ, Ju J. PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Scientific Reports* 2, Article number: 684, 2012. doi:10.1038/srep00684

<http://www.nature.com/srep/2012/120921/srep00684/full/srep00684.html>

「高速に大量のシーケンスデータはでてくるが・・・この解析がボトルネック」

- パーソナルゲノム時代が到来し、米国ではNIH・民間・医師会が共同し、個別化ゲノム医療の実施体制が急速に作られている。
- スパコン、大規模ストレージなどのリソースとデータ解析の人材が必要

2013年1月10日

米国National Institute of Health所長
Francis Collinsは、“NIH & Big Data”対応に
Associate Director for Data Scienceを設置し、人
材をリクルートすることを発表

“Tackling the Big Data Program”

- Centers of Excellence for Biomedical Big Data-

18



米国NIH所長
F. Collins

「私のがん」の黒幕たちを暴き出すために、なぜスパコンが必要なのか？

スーパーコンピュータで シーケンスのアセンブリとデータ解析



メディカルインフォマ ティクスの達人たち



検体

Ion Proton



半導体シーケンサー

6時間

正常組織 900億文字
がん組織 1200億文字
のデータ

Human Genome Center
Institute of Medical Science, University of Tokyo

スパコンシステム



1日~



親から受け
継いだゲノム

がんを発症
する鍵遺伝
子の変化

環境因子に
よる変化を
受けたゲノム

黒幕X

黒幕Z

*** 検査結果 時系列画面プリント ***

プリント時間: 2012. 4. 9 13:23:35 P. 1

患者番号: [REDACTED] 氏名: [REDACTED]

採取日	2011. 10. 19	2012. 01. 12	2012. 01. 18	2012. 04. 09
採取時間	00:00	00:00	00:00	00:00
総蛋白	6.9~8.4	7.1	7.0	7.8
アルブミン	3.9~5.2	4.1	3.8 L	4.4
CK	62~287	82	*** **	100
AST(GOT)	13~33	34 H	64 H	45 H
ALT(GPT)	8~42	39	72 H	49 H
LD	119~229	163	173	217
γGT	9~109	101	167 H	127 H
尿素窒素	8~21	13	13	13
クレアチン	0.6~1.1	0.9	0.8	0.9
eGFR	項目説明参照	71.2	76.6	68.2
尿酸	2.5~7.0	5.6	5.7	6.2
空腹時血糖	70~110(空腹時)	101	96	112
食事日時				
食事内容				
採血日時				
HbA1c (NGSP)	4.6~6.2	*** **	*** **	5.7
HbA1c (JDS)	4.3~5.8	5.5	5.5	5.3
中性脂肪	30~150	128	145	144
総コレステロール	122~240	191	200	226
HDLコレステロール	35~70	71 H	77 H	87 H
LDLコレステロール(直接法)	140未満	84	85	99
Na	139~146	142	142	141
K	3.7~4.8	4.1	4.1	4.1
Cl	101~109	104	104	102
Ca	8.7~10.1	8.9	8.8	9.4
無機リン	2.8~4.6	2.8	2.9	3.4
溶血(H)	項目説明参照	7	5	6
黄疸(I)	項目説明参照	1.5	1.8	2.5
乳び(T)	項目説明参照	1	1	1.8
血算				2
白血球数	3.4~9.2	5.9	4.8	6.1
赤血球数	4.00~5.66	4.70	4.64	4.86
ヘモグロビン	13.0~17.0	15.5	15.5	15.9
ヘマトクリット	38.2~50.8	45.4	45.3	47.1
MCV	85~102	96.6	97.6	96.9
MCH	28.4~34.6	33.0	33.4	32.7
MCHC	32.5~35.5	34.1	34.2	33.8
血小板数	141~327	178	155	162
比重	1.002~1.030	1.011	1.020	1.024
尿定性				1.013
pH	pH6(弱酸性)	5.5	5.5	5.5
糖	-	-	-	-
蛋白	-	-	-	-
潜血	-	-	-	+- H
ケトン体	-	-	-	-
ビリルビン	-	-	-	-
ウロビリノーゲン	1+	1+	1+	1+
亜硝酸塩	-	-	-	-
WBC定性	-	-	-	-
色調		淡黄色	淡黄色	黄色
濁度		-	-	淡黄色

パーソナルゲノム

親から受け継いだゲノム

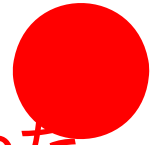
がんを発症する鍵遺伝子の変化

環境因子による変化を受けたゲノム

様々なリスクがわかる。治療方針が立つ。

これまでは「研究」のために行っていたシーケンスに対し、
がんや患者さんの全DNA情報を
シーケンスし、臨床的に翻訳・解
釈することが現実になった。

臨床シーケンスの時代に入った



全ゲノム臨床シーケンス(患者さんに返す)研究がはじまった。

- 2009-2011年: [“One In A Billion”](#)

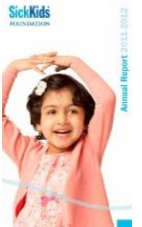
(<http://www.pulitzer.org/archives/9180>): 米国ウイスコンシン医科大学が世界で初めての[全遺伝子解析に基づく治療](#)を行った。そのレポート記事に2011年Pulitzer賞が与えられる。

<http://www.pulitzer.org/archives/9180>

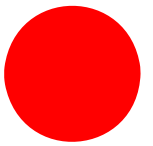
– 米国NIH所長Francis Collinsが[2012年予算要求演説](#)の結論の中で引用。

- 2011年12月6日: 米国NIH発表 “ゲノムシーケンスに基づく医療応用研究に\$4.16億ドル” (4年間で400億円)

- 2012年6月20日: トロントの [Hospital for Sick Children](#) が半導体シーケンサー(Ion Proton)を導入し、将来的には、1万人/年の規模で全ゲノム臨床シーケンスをすることを発表



臨床シーケンスの時代に入った



全ゲノム臨床シーケンス(患者さんに返す)研究がはじまった。

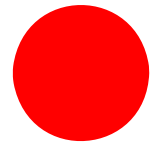
- 2012年2月2日:カナダ [Ontario Institute for Cancer Research](#) は、国際がんゲノムコンソーシアムの成果に基づき、がんの臨床シーケンスに基づく個別化医療の研究開始を発表
- 2012年:3500以上の単一遺伝子疾患が知られているが、新生児に対してそれらを一度に検索ために、全DNAシーケンスとそのデータ解析を約50時間で行い、診断をして治療方針を決定したという報告(University of Missouri-Kansas City: Science Translational Medicine. 4(154):154ra135).



Babies with genetic disorders can have their whole genome screened for mutations in just two days.

TAYLOR S. KENNEDY/GETTY IMAGES

臨床シーケンスの時代に入った



- 2013年1月28日～29日：米国NIHが主催したGenomic Medicine IVでは、“Physician Education in Genomics”がテーマで、循環器系、がん、小児科などの多数の医師会 (professional societies)が、ゲノム医療のための「医師の教育」に取り組んでいるかが報告された。
<http://www.genome.gov/27552294>
- 2013年9月には、Genomic Medicine Vで **“Unifying US Efforts for Coordinating Genomic Medicine Implementation”**がテーマ。
- 2014年1月には **“Genomic Medicine is Global”**をテーマに Genomic Medicine VIが開催される。国際連携が議論される予定。

「米国ではまさに明日の問題！」



Dr. Lukas Wartman
Washington Univ. St. Louis
白血病の研究者が白血病に
なり・・・

次世代シーケンサーとスパコンによるデータ解析で、腎癌の抗癌剤“スーテント”の有効性が示唆され、彼の白血病が寛解

"I was definitely scared. It was so unreal," said Dr. Wartman on first suspecting that he had leukemia, the very disease he had devoted his medical career to studying.

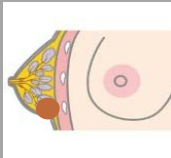
A drug that had been tested and approved only for advanced [kidney cancer](#). Dr. Wartman became the first person ever to take it for leukemia.

<http://www.nytimes.com/2012/07/08/health/in-gene-sequencing-treatment-for-leukemia-glimpses-of-the-future.html>

個別化医療のためのメディカルインフォマティクス(がんを例として)

全ゲノム解析をしていない従来の医療では

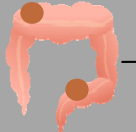
大腸がん患者さんでは、肺がんの原因遺伝子の変異は無視される



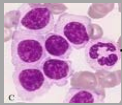
- Tamoxifen
- Anastrozole
- Doxorubicin
- Paclitaxel



- Cisplatin
- Etoposide
- Amrubicin
- Cyclophosphamide
- Doxorubicin



- 5FU
- UFT
- Oxaliplatin
- Irinotecan



- Idarubicin
- Daunorubicine
- Cytarabine

臓器別による抗癌剤の選択のみ



最先端シーケンサー



全ゲノム解析

AATGCCA
TCATGTC
GGTATCG
CAGC ...

薬剤応答システム解析



パーソナルゲノム・トランスクリプトーム・エピゲノム・
プロテオーム・メタボローム

個別化医療のための
メディカルインフォマティクス

翻訳・
解釈

バイオメディカル
ビッグデータ解析

大規模のゲノム及び遺伝子発現
データ、薬理・疾患情報のデータ
解析による個別化医療の基礎データ・知識ベース

全ゲノム・全トランスクリプトーム解析をすると

私のがんにぴったりあった副作用の無い抗癌剤と治療法がありました!



ゲノム変異別分類

抗がん剤・治療法

HER2 amplification

- Trastuzumab
- Lapatinib

EGFR mutation

- Gefitinib
- Erlotinib

ALK fusion

- Crizotinib

RET fusion

- Vandetanib

BRAF mutation

- Vemurafenib

PIK3CA mutation

- Temozolomide

BCR-ABL fusion
cKIT mutation

- Imatinib

Open[Geisinger Health System](#)

Implementing genomic medicine in the clinic: the future is here

Teri A. Manolio, MD, PhD¹, Rex L. Chisholm, PhD², Brad Ozenberger, PhD¹, Dan M. Roden, MD³,
Marc S. Williams, MD^{4,5}, Richard Wilson, PhD⁶, David Bick, MD⁷, Erwin P. Bottinger, MD⁸,

Although the potential for genomics to contribute to clinical care has long been anticipated, the pace of defining the risks and benefits of incorporating genomic findings into medical practice has been relatively slow. Several institutions have recently begun genomic medicine programs, encountering many of the same obstacles and developing the same solutions, often independently. Recognizing that successful early experiences can inform subsequent efforts, the National Human Genome Research Institute brought together a number of these groups to describe their ongoing projects and challenges, identify common infrastructure and research needs, and outline an implementation framework for investigating and introducing similar programs elsewhere. Chief among the challenges were limited evidence and consensus on which genomic variants were medically relevant; lack of reimbursement for genomically driven interventions; and burden to patients and clinicians of assaying, reporting, intervening, and following up genomic findings. Key infrastructure needs included an openly accessible knowledge base capturing sequence variants and their phenotypic associations and a framework for defining and cataloging clinically actionable variants. Multiple institutions are actively engaged in using genomic information in clinical care. Much of this work is being done in isolation and would benefit from more structured collaboration and sharing of best practices. *Genet Med* advance online publication 10 January 2013 *Genetics in Medicine* (2012); doi:10.1038/gim.2012.157.



Mayo Clinic



Richard Weinshilboum教授(薬理ゲノム)

- メイヨーは5年以内に10万人の全ゲノム情報のデータを取り、電子カルテとリンクさせる
- 専用のコンピュータ棟も建築中
(2013年1月の情報)

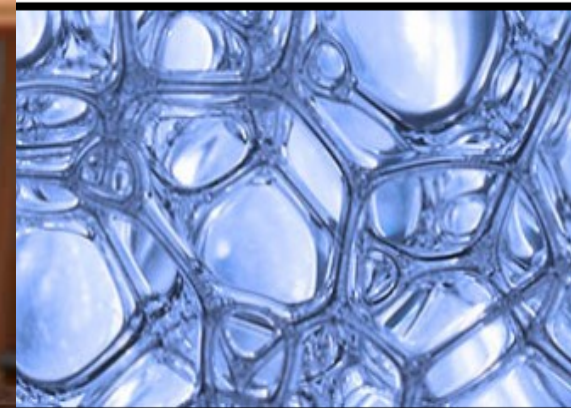
メイヨークリニックのあるミネソタ州ロチェスターは人口10万人の町で、メイヨークリニック関係の雇用が3万5千人、医師がレジデントとを含めると5000人。入院ベッドは2000床で、年間延べ患者数は100万人、この内隣接する地域が50%、その他の米国からが45%、海外から5%(5万人)。ロチェスター空港はボーディングブリッジは2つしかありませんが、著名人がメイヨーを受診するため、プライベートジェットで乗り付けるそうです。完全な医療都市です。アリゾナとフロリダにも病院を持っています。

シカゴ大学 [Grossman](#) 教授のブログ

100万人ゲノムが医療を変える

- **Managing and Analyzing 1,000,000 Genomes**: September 18, 2012 by Robert Grossman
- Recently, we have been thinking about what you might call the [Million Genome Challenge](#). Over the next several years, the National Cancer Institute, and perhaps other organizations, will sequence a million genomes and use this data to increase our understanding of biological pathways and of genomic variation across individuals. [With this knowledge, we can begin to stratify diseases, leading to precision diagnosis and precision treatment that is personalized for individual patients.](#)
- [The numbers associated with a million cancer genomes are worth thinking about.](#) The whole genome data for a tumor and a matching normal tissue sample require about 1 TB. Thus, one million genomes require about 1,000,000 TB. This is 1,000 PB or 1 EB. Compressing the data might reduce the data by about a factor of 10. Throwing away the alignment data and retaining only the variation data would reduce the data by about a factor of about 100. Assuming it costs about \$1,000 to sequence each whole genome, the project as a whole requires about \$1B for the sequencing. It might require another \$1B for the infrastructure and analysis. Although obviously a large project, [a project like this is likely to fundamentally alter the way we understand and treat diseases.](#)

Robert Grossman



[Group](#) [Software](#) [Blog](#) [Contacting](#)

[← Why is it still so hard to analyze remote and distributed data?](#)

[Datascope for the Long Tail of Science →](#)

Managing and Analyzing 1,000,000 Genomes

Posted on [September 18, 2012](#) by [Robert Grossman](#)

I gave a talk last week at [XLDB 2012](#) about [Bionimbus](#), which is cloud based system for managing, analyzing, transporting, and sharing large genomics datasets in a secure and compliant fashion. Bionimbus was developed at the Institute for Genomics and Systems Biology ([IGSB](#)) and is used by IGSB and some of their collaborators to manage and analyze their next gen sequencing data.

Knocking on the clinic door

Nature Biotechnology **30**, 1009 (2012) | doi:10.1038/nbt.2428

Published online 08 November 2012

[Download PDF](#)[Citation](#)[Reprints](#)[Rights & permissions](#)

High-throughput sequencing for clinical purposes faces technical and quality challenges, but it's worth it.

The \$3-billion price tag for the first human genome would now buy not one but a million human genome sequences, each completed in just a few weeks. Personal genome sequencing is becoming a reality, and targeted or whole exome sequencing is being explored to facilitate diagnosis and guide treatment, in some conditions and for some patients. The problem is that extracting clinically actionable information from genome data is currently hit or miss, time intensive and dependent on access to knowledgeable specialists. What's more, much of the IT infrastructure and decision support systems necessary to deliver genome information to physicians has yet to be put in place.

我が国の風景

パーソナルゲノムに基づく個別化医療の推進(これからの医科研5年間)

がん・感染症
患者



2001年より遺伝子カウンセリングと遺伝子検査体制が整備されている。

医科研附属病院

お金はないですが、
やります！



遺伝カウンセリング

検体

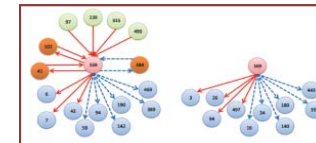
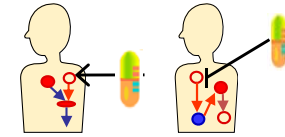
超並列シーケンサー
2013年ヒトゲノムは10万円以下で解析

スパコンによるデータ解析
225 TFLOPS計算ノード

全DNA解析
トランスクリプトーム解析
エピゲノム解析

パーソナルオミクス情報
1人当たり0.5TBを超えるデータ

医療介入の予測

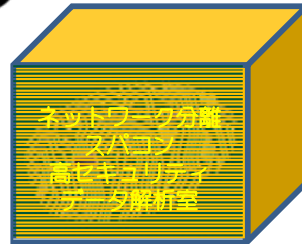


- 最適な治療法・治療薬量の決定
- 副作用の回避
- 個別化検診・サーベイランスの実施
- 体質に応じた予防法

Human Genome Center
Institute of Medical Science, University of Tokyo



4.4PBストレージ
3PB Lustre File System



先端医療研究センター・医科研附属病院

ヒトゲノム解析センター

スパコン: 国際がんゲノムコンソーシアムを初めとするゲノムシーケンシングデータ解析の体制が整っている。

血液がん: 年間400件以上の多様な遺伝子検査と診療の実績
大腸がん: 100症例以上の遺伝性大腸がんの遺伝子診断と診療

Personalized
Medicine
Team

ヒトゲノム解析センター

- 計算能力: 225 TFLOPS (ピーク性能)
- ストレージ: Lustre 高速ストレージ(3PB) + ニアライン(2PB)
- リース: 6年リースで60億円弱 (2009.1~2014.12)
- 消費電力: 実測値550KW~800KW (ピーク値1200KW)
- 多様なライフサイエンスアプリに対応: **22,704コア**

	ノード数	CPU	メモリ/ノード	総メモリ	総コア数
Shirokane 2(分散メモリサーバ)	504	CPU:AMD Opteron 6276 2.3GHz (16cores) × 2	32GB	16.128TB	16,128
Shirokane2(分散メモリサーバ)	12	CPU:Intel Xeon X5675 3.06GHz (6cores) × 2	144GB	1.728TB	144
Shirokane2(共有メモリサーバ)	1	CPU:Intel Xeon E7 8837 2.66GHz (8cores) × 16	2,000GB	2TB	128
Shirokane1(分散メモリサーバ)	764	CPU:Intel Quad Core Xeon E5450 3.0GHz (4cores) × 2	32GB	24.448TB	6,112
Shirokane1(分散メモリサーバ)	12	CPU:AMD Quad Core Opteron 8356 2.3GHz (4cores) × 4	128GB	1.536TB	192

SHIROKANE システム全体構成図

Shirokane1

Shirokane2

3階

共有メモリ型サーバ

分散メモリ型サーバタイプA

分散メモリ型サーバタイプB

分散メモリ型サーバタイプB

分散メモリ型サーバタイプA

CPU: Intel Xeon E7 8837 2.66GHz (8cores) × 16
メモリ: 2TB

12ノード
CPU: AMD Quad Core Opteron 8356 2.3GHz (4cores) × 4
メモリ: 128GB

764ノード
CPU: Intel Quad Core Xeon E5450 3.0GHz (4cores) × 2
メモリ: 32GB

504ノード:
CPU: AMD Opteron 6276 2.3GHz (16cores) × 2
メモリ: 32GB

12ノード
CPU: Intel Xeon X5675 3.06GHz (6cores) × 2
メモリ: 144GB

1階

容量 500TB

データバックアップサーバ

高速データ通信ネットワーク

高速データ通信ネットワーク

容量 500TB

データバックアップサーバ

容量 1PB

高速ディスクアレイ

容量 500TB

DBサーバ

容量 2.3PB

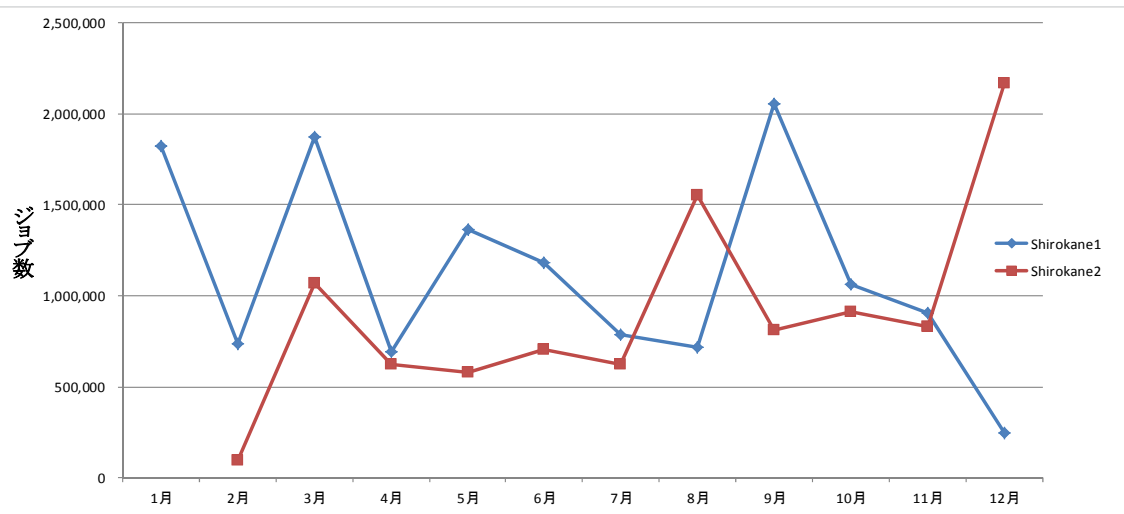
高速ディスクアレイ

パーソナルゲノム時代の 基盤ファイルシステム

ヒトゲノム解析センターの場合
ライフサイエンス系利用に特化したコンピュータ

メタデータアクセス

- ジョブの数、ジョブの入出力ファイルの数が多い
- 1ジョブセットで数万ファイルが作成・参照されることも多い



	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
Shirokane1	1,822,799	737,649	1,876,145	688,983	1,362,230	1,182,873	789,337	714,301	2,058,491	1,062,914	906,922	244,279
Shirokane2		97,451	1,071,403	623,953	577,121	706,634	621,990	1,553,523	809,651	910,357	829,960	2,166,036
	ジョブ数(2012年合計)		月平均									
Shirokane1	13,446,923		1,120,577									
Shirokane2	9,968,079		906,189									

ヒトゲノム解析センター	2012年度実績
Shirokane1	13,446,923 ジョブ
Shirokane2	9,968,079 ジョブ
合計	23,415,102 ジョブ

比較のためにつけております

東大情報基盤センター	2012年年度実績
FX10	322,555 ジョブ (4月-12月)
HA8000	91,979 ジョブ
SRI16000	58,041 ジョブ

困難はいろいろありますが

だれもが保険証を持っているように自分のDNA情報をもって医療を受ける時代がやがて来る。

DNA情報に基づいた個別化医療

遺伝子 医療革命

ゲノム科学がわたしたちを変える
The Language of Life

DNA and the Revolution in Personalized Medicine

フランシス・S・コリンズ Francis S. Collins 矢野真千子訳

史上最強のアイスホッケー選手とだれもが認めるウェイン・グレッツキーは、彼の最初にして最高のコーチである父親ウォルターから、「バックが行きそうなところに行け」というシンプルな教えを受けていたという。私たちはみな、人生というアイスホッケー試合のスケーターだ。チームのメンバーと協力しながら、転ばないように注意しながら、そしてできることなら数回ゴールを決めながら、リングの上を巧みに動き回ろうとしている。そのためには、バックを追いかけるだけではだめだ。バックが行きそうなところに行っておかなければならない。あなたのDNAの二重らせん、あなたの生命の言語は、あなただけの医学書だ。その読み方を学び、人生に活かすことができるのは、あなたしかいない。

Francis S. Collins著(米国National Institute of Health所長)
(邦訳 NHK出版)

