

平成 27 年度
「個別化・予防医療を支援する統合計算生命科学」
成果報告書

平成 28 年 5 月 27 日
国立大学法人東京大学
宮野 悟

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、国立大学法人東京大学、が実施した平成27年度「個別化・予防医療を支援する統合計算生命科学」の成果を取りまとめたものです。

内容

1. 委託業務の題目	1
2. 実施機関（代表機関）	1
3. 委託業務の目的	1
4. 平成27年度（報告年度）の実施内容	2
4-1. 実施計画	2
(1) 大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明（サブ課題A）	2
(2) データ同化生体シミュレーションによる個別化医療支援（サブ課題B）	2
(3) 心臓シミュレーションと分子シミュレーションの融合による基礎医学と臨床医学の架橋（サブ課題C）	3
(4) プロジェクトの総合的推進	3
4-2. 実施内容（成果）	3
(1) 大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明（サブ課題A）	3
(2) データ同化生体シミュレーションによる個別化医療支援（サブ課題B）	13
(3) 心臓シミュレーションと分子シミュレーションの融合による基礎医学と臨床医学の架橋（サブ課題C）	23
(4) プロジェクトの総合的推進	27
4-3. 実施体制	27

1. 委託業務の題目

「個別化・予防医療を支援する統合計算生命科学」

2. 実施機関（代表機関）

代 表 機 関	機関名		国立大学法人東京大学			
	所在地		〒113-8654 東京都文京区本郷七丁目3番1号			
	課題 責任者	ふりがな	みやの さとる	生年	西暦 1954 年 12 月 5 日 (60 歳)	
		氏 名	宮野 悟	月 日	※2015 年 4 月 1 日現在	
		所属部署名	医科学研究所		役職	教授
		連絡先	Tel. 03-5449-5615		Fax. 03-5449-5442	
	E-mail:miyanolab-jimu@edelweiss.hgc.jp, miyano@ims.u-tokyo.ac.jp					
	事務 連絡 担当者	ふりがな	さいとう としや			
		氏 名	齋藤 利弥			
		所属部署名	医科学研究所 研究支援課 外部資金戦略チーム		役職	主任
連絡先		Tel. 03-5449-5137		Fax. 03 - 6409 - 2017		
	E-mail t-gshikin@ims.u-tokyo.ac.jp					

3. 委託業務の目的

病気は、臓器群の変調という現象として現れるが、その背景には生命の設計図とも呼ばれるゲノムがあり、オミクスとよばれるエピゲノム、RNA、タンパク質など多彩な分子が細胞を制御・構成している。また、細胞には環境や加齢により長い時間をかけて変化していく個々人で異なる細胞コンテキストがあり、そのもとで構成されている臓器の状態も多様である。さらに、その繋がりは人智を超えた複雑さを有している。その理解には、画像や生理データなどを含む高精度臨床データとともに環境・生体時空間的にゲノムから全身を捉える必要がある。現在、ゲノム解析技術の劇的な革新と高精度計測機器の急速な発展は、健康・医療ビッグデータを生みだそうとしており、この傾向は急激に加速している。こうした背景のもと、本研究の目的は、ポスト「京」によって初めて実現できる「情報の技術」と「物理の原理」の融合により、がんをはじめとして全脳・循環・代謝系など、全身の疾患に対して、ビッグデータを活用し、高度の生体階層統合シミュレーションに個体データを同化させる技術、並びに、ライフサイエンスにおいてかつてない大規模なデータ解析技術を開発・応用することにより、また、大規模データに基づくアプローチと並行して、分子細胞レベルの研究と臓器個体レベルの研究を融合させ、ミクロとマクロのメカニクスとを関連させて定量的とらえたシミュレーションモデルを構築することにより、病態の理解と効果的な治療法の探索を行い、その成果を個別化・予防医療へ返す基盤となる統合計算生命科学を確立することを目的とする。

このため、国立大学法人東京大学を中核機関として、分担機関である国立大学法人京都大学、国立大学法人大阪大学、株式会社UT-Heart研究所、学校法人自治医科大学、国立大学法人岡山大学と密接に連携し、研究開発を実施する。

4. 平成27年度（報告年度）の実施内容

4-1. 実施計画

平成27年度は、平成26年度に引き続き、調査研究・準備研究フェーズとして、以下に示す重点課題②に関するサブ課題（A～C）及びその実施内容について、開発計画（研究開発内容、目標・期待される成果、実施体制、必要計算資源、工程表、所要経費等）の詳細の策定をする。また、策定した開発計画に基づき、アプリケーション開発の本格実施に向けた準備（実施体制構築の着手、コデザインの先行作業等）を継続して進める。また、サブ課題の内容が指摘事項に合致しているか継続して検討する。本業務を遂行するためにノートパソコン一台を計上している。

（1）大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明（サブ課題 A）

がんの個別化医療・予防の実現のためには、個々人のがんの相違、1つのがんの内部での多様性・起源を、全ゲノムシーケンス、RNAシーケンス、メチル化シーケンス、単一細胞シーケンス、さらには血漿をソースとしたDNAシーケンスから得られる大量シーケンスデータ解析を用いて明らかにする必要がある。本サブ課題では、現実のがん試料のデータに即して、「京」コンピュータ、及びポスト「京」の特段の計算能力を駆使することにより、こうした超大規模データ解析に必要となる現実的な計算機ソリューションを提供することにより、ゲノムシーケンスに基づいた個別化・予防医療の実現を強力に支援するための統合計算生命科学を創設する。研究統括、大規模データ解析、システム統合解析、ソフトウェア開発、がん発生進化シミュレーションを実施する。

平成27年度は、「京」上での大規模シーケンスデータ解析環境の整備とがん研究のビッグデータ化状況について調査する。

①大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明

大規模がん試料に関する全ゲノムシーケンス、RNAシーケンス、メチル化シーケンス、単一細胞シーケンス、さらには血漿をソースとしたDNAシーケンスから得られる大量シーケンス解析、及びデータ解析を実施する。

（2）データ同化生体シミュレーションによる個別化医療支援（サブ課題 B）

全脳循環代謝シミュレータの開発を実施する。データ同化イメージベーストシミュレーションによる脳動脈瘤の診断治療支援の研究開発を実施する。

平成27年度は、3次元内部構造顕微鏡を用いてマウス脳血管の形態計測を行い、全脳循環代謝シミュレータ開発に向けた脳血管モデルを構築する。また、個別化医療支援に向けた以下の大規模シミュレータの開発計画の詳細を策定し、アプリケーションの開発準備を行う。

①実測データを用いた循環器系疾患のシミュレーション

②微小循環遊走細胞解析プラットフォームの開発

③超大規模脊髄神経・筋骨格系モデルの開発と動態シミュレーション

④医療用画像の統計的機械学習による形態と疾患との関連性の発見

(3) 心臓シミュレーションと分子シミュレーションの融合による基礎医学と臨床医学の架橋（サブ課題 C）

心臓シミュレータUT-Heartと分子シミュレータCafeMolを融合させることにより、ミクロ・マクロ間の相互作用により病態が進行する心不全の解明と最適治療を可能とする世界でも例を見ないマルチスケール心臓シミュレーションを実現する。医療への実用化を展開する。

平成27年度は共同研究を行う医学系のグループと共に当初策定した開発計画を見直し、アプリケーション開発の本格実施に向けた準備を行う。

- ①心不全を中心とする臨床情報とりまとめ、提供、及び、シミュレーション結果の検証
- ②小児先天性心疾患を中心とする臨床情報とりまとめ、提供、及びシミュレーション結果の検証

(4) プロジェクトの総合的推進

プロジェクト全体の連携を密としつつ円滑に運営していくため、運営委員会や研究会の開催等、参画各機関の連携・調整にあたる。

特に、プロジェクト全体の進捗状況を確認しつつ計画の合理化を検討し、必要に応じて調査或いは外部有識者を招聘して意見を聞くなど、プロジェクトの推進に資する。

プロジェクトで得られた成果については、積極的に公表し、今後の展開に資するが、UT-Heartのコードについては知財の観点から公開は行わない。

4-2. 実施内容（成果）

平成27年度は、調査研究・準備研究フェーズとして、以下の内容を実施した。特に、(1)を中心にして、コデザインのための既存開発アプリケーションソフトウェアの性能評価、及びボトルネックの同定を行った。

(1) 大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明（サブ課題 A）

サブ課題 A は、大規模がん試料に関する全ゲノムシーケンス、RNA シーケンス、メチル化シーケンス、単一細胞シーケンス、さらには血漿をソースとした DNA シーケンスから得られる大量シーケンス解析、及びデータ解析を実施することを目標として、調査研究・準備研究フェーズとして、その実施内容について、開発計画（研究開発内容、目標・期待される成果、実施体制、必要計算資源、工程表、所要経費等）の詳細の策定を行った。また、策定した開発計画に基づき、アプリケーション開発の本格実施に向けた準備（実施体制構築の着手、コデザインの作業等）を行った。具体的には、がん試料のデータに即して、超大規模データ解析に必要となる現実的な計算機ソリューションを提供することにより、ゲノムシーケンスに基づいた個別化・予防医療の実現を強力に支援するための統合計算生命科学を創設するために、「京」上での大規模シーケンスデータ解析環境の整備とがん研究のビッグデータ化状況について調査した。以下にその詳細についてのべる。この調査研究のなかで、すでに画期的な成果が出始めている。

1. 大規模データ解析パイプラインのヒトゲノム解析センタースーパーコンピュータ上での整備と実証
がんの変異には、がんの種類に応じて明らかな傾向があることが知られている。例として、喫煙歴の

ある肺がんについては、タバコに含まれる化学物質によりもたらされる C>A の変異が多く観察されることが、また皮膚がんにおいては、紫外線による C>T、CC>TT の変異が多く観察されることが知られていた。近年目覚ましく発展しているシーケンス技術の発展により、新規がん原因遺伝子の同定だけでなく、個々のがんゲノムにおける変異のプロファイルの違いをこれまでにない精度で検出することがある程度可能になっていることが調査結果確認できた (Nik-Zainal et al., Cell, 2012, Alexandrov et al. Nature, 2013)。そのため、新たな変異パターンの発見、またそれに付随する発がん物質を同定、さらにはそれを新規発がん物質の発見や評価へとつなげ、がんの予防法へと持っていくなかで、がんの個性と時間的・空間的多様性・起源の解明することが極めて重要な事項であることが本調査研究のなかで判明した。後述するが、一部、その成果が見えてきた。一方で、がんゲノムシーケンス解析による大量の変異データから、特徴的なパターンを抽出するために、新たな情報学的・統計学的手法の開発が重要であることも判明した。現在、主流と考えられている NMF (nonnegative matrix factorization) による方法論には、「組み合わせ爆発の問題から、考慮する因子を増やそうとすると、パラメータ数が指数的に増大し、推定が著しく不安定になってしまう」ということなどの問題点があることが判明した。この問題点を、probabilistic mutation signature (pmsignature、<https://github.com/friendlws/pmsignature>) という新しい手法を開発することによって解決した。ソフトウェアは github で公開した (図 1)。この方法は、「条件付き独立性を仮定したモデリングにより、変異パターンの因子数を増やしても、首尾よく推定が可能である」また、「提案モデルは機械学習分野で文書分類に利用されるトピックモデル (Blei et al. JMLR, 2003) と類似したモデルとなっており、過去にこれらの分野で蓄積されてきた膨大な知見を利用することが出来る」などの特徴を備えている [1]。そして、これまで、サブ課題 A 課題分担者である小川誠司 (京都大学医学研究科) と宮野悟 (東京大学医科学研究所) とは、6 年にわたる共同研究において、がんゲノムシーケンスデータをスーパーコンピュータ上で効率的に解析するプラットフォーム Genomon (<http://genomon.hgc.jp/exome/en/>) の開発を通じ、多数の新規がん原因遺伝子の発見に貢献してきた実績を有している (2011-2016 現在まで : New Eng J Med (1), Nature (2), Science (1), Nature Genetics (9) など)。エクソーム解析の限界が見えてくる中、全ゲノム解析の必要性がますます高まっている。そのため、本調査・準備研究においては、これまで Genomon に用いられていたプログラムを包括的に改善し、高度に効率的に全ゲノム解析を可能とする新規プラットフォーム Genomon2 (<http://genomon-project.github.io/GenomonPages/>) の開発を行った (図 2)。Genomon2 は、偽陽性の変異を効果的にフィルターするプログラム GenomonMutationFilter (<https://github.com/Genomon-Project/GenomonMutationFilter>)、多数のコントロール検体を効率的に利用して偽陽性を除くプログラム、EBFilter (<https://github.com/Genomon-Project/EBFilter>) を備え、特に複雑な配列パターンを有する全ゲノム領域においても高度に正確な変異検出を可能としている。また、GenomonSV (<https://github.com/Genomon-Project/GenomonSV>) により、大規模な構造変異だけではなく、数十塩基から数百塩基の中程度のサイズの構造変異の検出が可能である。さらに、ruffuse, drmma など最新のジョブマネジメントツールを効率的に利用した GenomonPipeline (<https://github.com/Genomon-Project/GenomonPipeline>) により、多数検体の効率的な解析を容易にしている。

ヒトゲノム解析センターでは Shirokane3 が、平成 27 年 4 月より稼働し始め、Genomon2 は Shirokane3 にスムーズに移植され、現在多くの共同研究者に利用されるようになり高負荷の下ではあるが何とか世

界トップレベルの成果を出している (Nature Genetics 2015 [3])。今後の大規模データに対応するために、「京」コンピュータ及びポスト「京」の利用が不可欠であり、まずは「京」コンピュータへの移植作業、及びがんの統合的理解のための応用が要である。

ネットワーク解析のための SiGN シリーズ等の「京」への移植はスムーズに行われた。これまで最適構造探索は他グループによる 33 変数が最大であったが、「京」コンピュータで採用されているトーラスインターコネクトネットワークに最適化された新しい並列アルゴリズムを考案し、離散モデルを用いて 36 変数の構造探索の実証に成功し、最適ベジアンネットワーク推定において世界記録を達成した。京互換機である FX10 4800 ノードによる成果で 36 変数 200 サンプルのデータに対し 11 時間 38 分 21 秒で全探索を完了した。

しかし、後述のように、Genomon シリーズはデータ分散の問題やストレージの運用方針による研究へ障害、ストレージの I/O の問題、コンパイラのバグなど、困難な問題も同定することも本準備・調査研究の中で行うことができた。これまで、最も負荷の大きい Genomon-fusion の「京」への効率的移植に成功したが[2]、今後は、世界トップの成果を出し続けている Genomon2 の「京」への移植に精力をつぎ込み、また「京」の運用上の協力を得ることが社会的理解をさらに得るためには不可欠であるとの結論に至った。

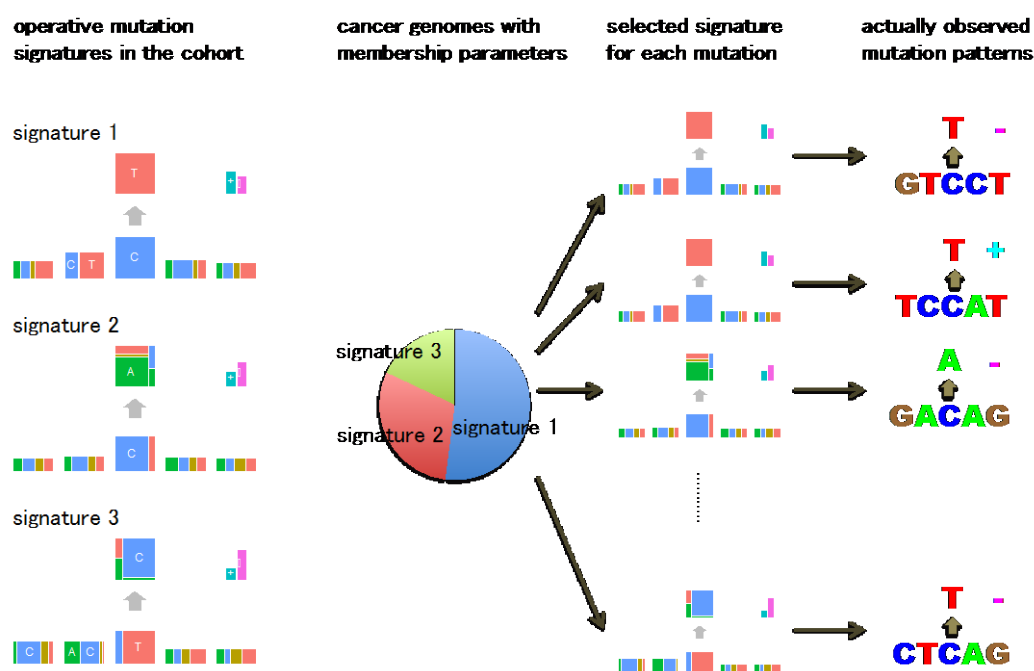


図 1:新手法の概念図 (上手) (Shiraishi et al. PLoS Genetics 2015). Genomon2 の DNA 解析のスキーマ図 (下図): 配列のアラインメント、点変異・欠失・挿入、構造変異の検出を行う

- [1] Shiraishi Y, Tremmel G, Miyano S, Stephens M. A simple model-based approach to inferring and visualizing cancer mutation signatures. *PLoS Genet.* 11(12):e1005657, 2015.
- [2] Ito S, Shiraishi Y, Shimamura T, Chiba K, Miyano S. High performance computing of a fusion gene detection pipeline on the K computer. *BIBM 2015.* 1441-1447, 2015.
- [3] Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga J, Totoki Y, Chiba K,

Sato-Otsubo A, Nagae G, Ishii R, Muto S, Kotani S, Watatani Y, Takeda J, Sanada M, Tanaka H, Suzuki H, Sato Y, Shiozawa Y, Yoshizato T, Yoshida K, Makishima H, Iwanaga M, Ma G, Nosaka K, Hishizawa M, Itonaga H, Imaizumi Y, Munakata W, Ogasawara H, Sato T, Sasai K, Muramoto K, Penova M, Kawaguchi T, Nakamura H, Hama N, Shide K, Kubuki Y, Hidaka T, Kameda T, Nakamaki T, Ishiyama K, Miyawaki S, Yoon SS, Tobinai K, Miyazaki Y, Takaori-Kondo A, Matsuda F, Takeuchi K, Nureki O, Aburatani H, Watanabe T, Shibata T, Matsuoka M, Miyano S, Shimoda K, Ogawa S. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nature Genetics*. 47(11):1304–1315, 2015.

2. Genomon-fusion for K の性能予測/高速化および多数検体解析実証、及び Genomon2 の移植設計

2.1 高速化および性能予測

重点課題2のターゲットアプリケーションである Genomon2 のプロトタイプとして、Genomon-fusion for K computer（以下、GFK）の高速化、および性能評価を実施した。Genomon-fusion は 東京大学 医科学研究所 ヒトゲノム解析センター宮野悟研究室と京都大学大学院医学研究科腫瘍生物学講座小川誠司研究室が共同して開発したものであり、ヒトゲノム解析センターに実装されている。Genomon-fusion は Whole Transcriptome シーケンスの結果である FASTQ ファイルのマッピング、融合遺伝子を検出し、融合遺伝子の候補の一覧を出力する様々なオープンソフトウェアを組み合わせたパイプラインと呼ばれるソフトである。ライセンスの元で自由に使用・改変・再配布が可能となっている。GFK は計算負荷及び I/O 負荷の大きく、また Genomon2 と同様のパイプラインソフトウェアであり、複数のソフトウェアを組み合わせて使う点や多数の中間ファイルによる入出力を伴う点など共通点が多いためターゲットアプリケーションとした。

GFK の高速化に関しては、blat, Perl スクリプト等の個別ソフトウェア/スクリプト単位でのプロファイリング結果からコストの高い順に詳細解析を行い、チューニングを実施した。具体的には、GFK 全体で最も計算コストのかかる blat、特に GFKalign 部分における処理内容を、詳細プロファイラ等を用いて解析し、OpenMP のスレッド並列効率の上昇（図3）、GFK ワークフローにおける非依存性処理のタスク並列化による実行時間短縮、perl スクリプトの条件分岐判定と後続処理の効率化などを行った。

以上のように高速化された GFK をベースに、性能評価ツールとサンプルデータのテスト計算から、「京」コンピュータ及びポスト「京」における性能予測および全系利用時の 1 日当たりの解析検体数の最大値を見積もった。性能評価方法としては、GFKalign、GFKdedup、GFKdetect のそれぞれについて個別に評価することとし、計算負荷最大の blat については詳細プロファイル計測を実施し、その結果を性能評価ツールにかけることでポスト「京」での消費電力および実行時間等を見積もった。また、I/O 律速となる Perl スクリプトについては、ポスト「京」での I/O 性能が「京」コンピュータに準じるものと仮定し、「京」での実測値を用いた。それ以外の小規模プログラム/スクリプトに関しては FX100 での実測値に対して、ポスト「京」との駆動周波数による線形補足値を用いている。その結果、チューニング前の GFK が約 3,000 検体/日だったのに対し、チューニング後は 5,356 検体/日と約 11 倍の性能向上を実現した（表1）。性能予測の詳細に関しては、理化学研究所 計算科学研究機構の詳細設計報告書および HPCS2016 採択論文[4]を参照されたい。

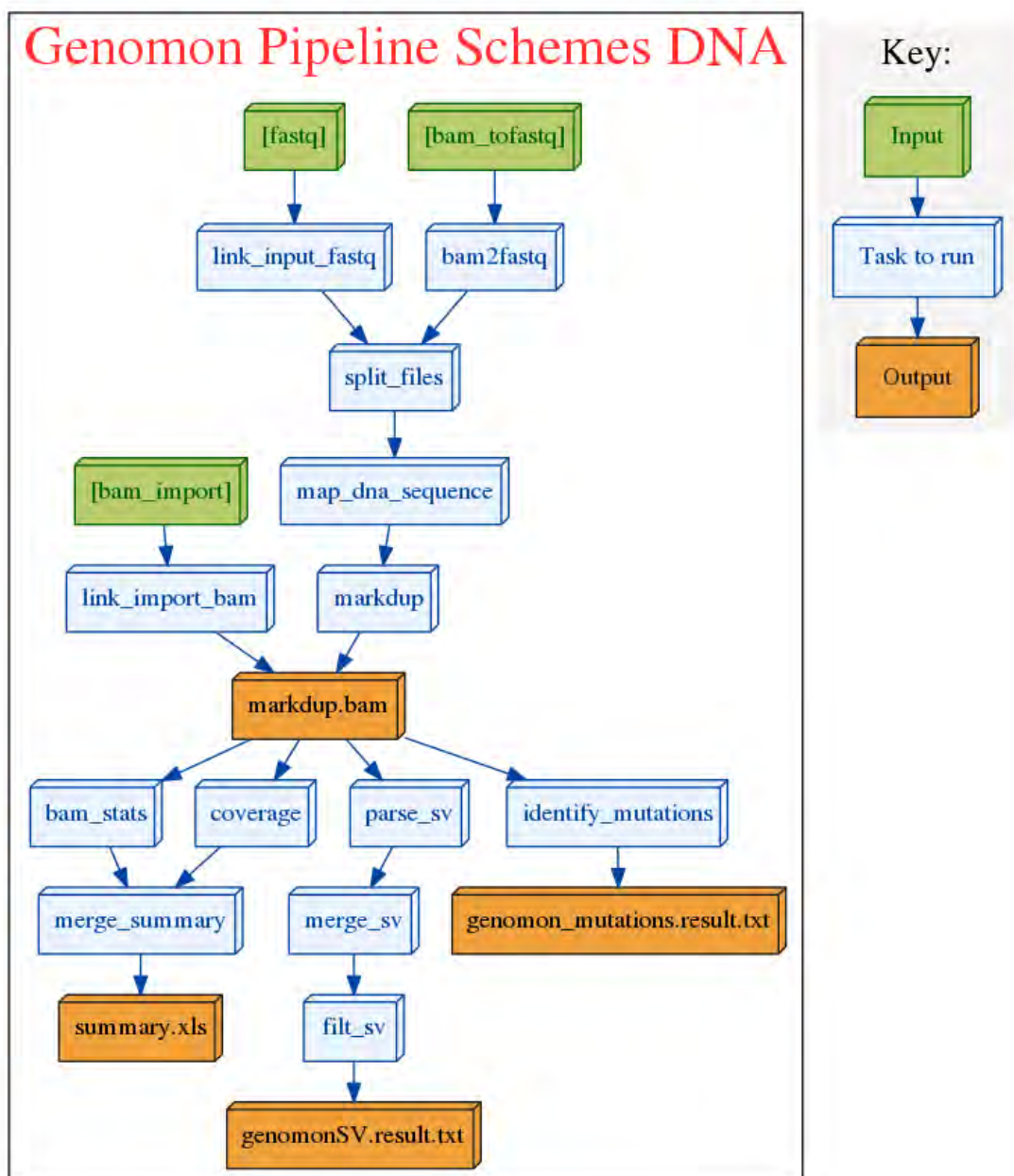
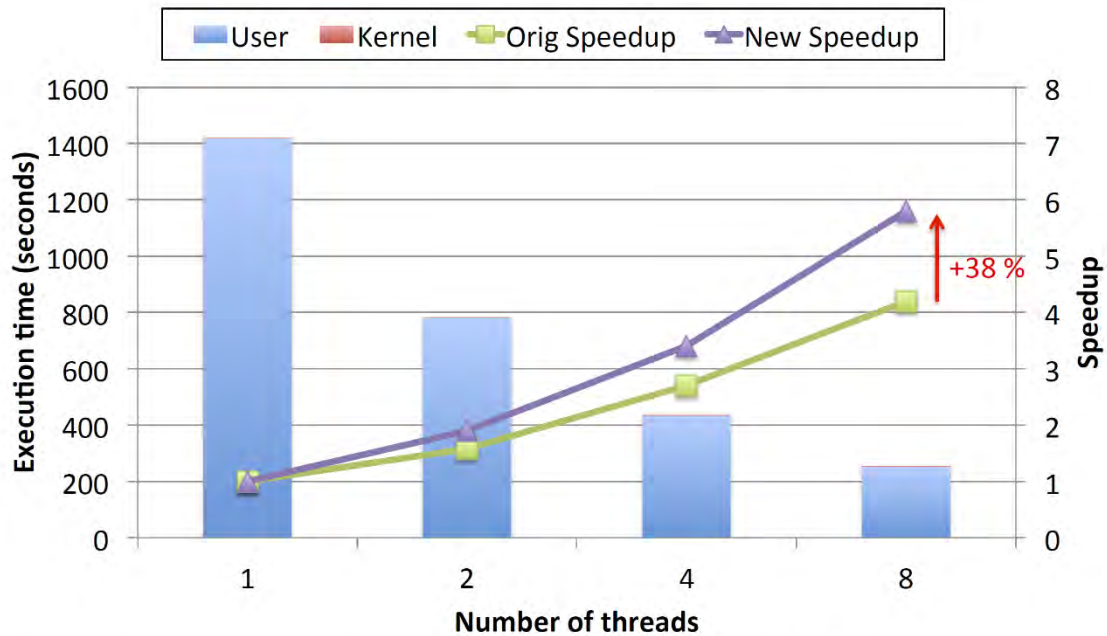


図 2： Genomon2 のパイプラインスキームの簡略化説明

blat – OpenMP with schedule(dynamic)



#pragma omp for -> #pragma omp for schedule(dynamic, 10)

図3：OpenMP 化 blat の動的スケジューリング導入による効率化

表1：「京」全ノードを用いた複数検体パイプライン実行[2]

	オリジナル GFK	チューニング後
1 ステージ実行時間 (秒)	24, 912	2, 387
ノード数/パイプライン	93	558
パイプライン数	891	148
解析検体数/日	3, 090	5366

[4] 鈴木惣一朗, 伊東聡, 池田奈生, Balazs Gerofi, 三吉郁夫, 丸山直也, 滝澤真一朗, 村瀬洋介, 石川裕, 宮野悟, ヒトゲノム解析プログラム Genomon-fusion の大規模スーパーコンピュータ向けの最適

2.2 多数検体解析実証

2.1での性能予測は主に1検体またはその一部分を用いたテストデータによるテスト計算結果からの性能予測である。しかし、オミックスデータは検体ごとのばらつき（データサイズやデータの質）が大きく、また、多数検体の実行による I/O 競合などの問題も懸念される。そこで、オリジナル GFK を用いて実際に多数検体の全解析を実施 [2] し、2.1 の予測およびテストデータの妥当性について検証した。

解析に利用したデータは CCLE (Cancer Cell Line Encyclopedia: <http://www.broadinstitute.org/ccle/home>) の RNA-seq データ 669 サンプルである。解析結果の概要を表2に示す。GFKalign、GFKdedup、GFKdetect の各パートごとの1検体当たりの平均消費リソース（ノード時間）との比較から、ベンチマークデータおよびそこから導かれる性能予測値の妥当性が確認できた。

表2：GFK を用いた多数検体解析の概要

	Elapse time (hour)	# of cores	Time / sample (seconds)	Benchmark (seconds)
GFKalign	875122.2	1,574,540	8,000	5,714
GFKdedup	4512.8	669	24,279	24,915
GFKdetect	53143.0	29,436	6,499	5,742
total	932778.0	1,604,645		

2.3 Genomon2 の移植設計

重点課題2におけるターゲットアプリケーションは全ゲノム (Whole Genome Sequence:WGS) 解析を目標としている。GFK の対象とするデータは RNA-seq であり、ポスト「京」完成時の本解析には耐えられない。平成27年10月末に WGS 解析用の新しいパイプライン「Genomon2」のアルファ版を内部リリースし、Genomon2 の京への移植作業を開始した。Genomon2 は WGS/Exome/RNA-seq 等のマルチデータ対応、点変異/融合遺伝子検出/構造変異検出等の複数種類解析対応の包括的なオミックスデータ解析パイプラインとなっている。

平成27年度では、Genomon2 について

- ・ワークフローの詳細解析
- ・各ソフトウェア/スクリプトの動作確認
- ・1検体解析と実施するための Sparc64VII fx 環境 (FX10) への移植

を実施した。ワークフローの解析は、WGS 解析がターゲットであるため DNA データ・mutation のフローに関して、各ステップにおけるインプット/アウトプットファイルの詳細、条件分岐、使用ソフト等をすべて洗い出している。この結果に基づき、FX10 環境への移殖を実施、各ソフトウェアの動作を確認し、不具合のあるソフトウェアについてはパッチの開発を実施した。

パイプラインタスク	利用ツールなど	標準入力サンプルに対する最大要求メモリ	RUFFUSパイプラインのマルチプロセス ※DRMAAを使うとマルチプロセス分だけのバッチジョブがサブミットされる ※サブミットされるジョブ数上限は制限可能(デフォルト100)		
			標準入力サンプル	1検体 入力bam 500GiB (350-450GiB)	京での実行目標 100検体 入力bam 500GiB
Task1A	bamtofastq	58.4 MiB	2(normal/tumor)	2	200
Task3A	split	2.9 MiB	2(normal/tumor)	2	200
Task4A	bwa/bamsort	5343.1 MiB	2(splitで分割する数に依存)	854×2(normal/tumor) = 1708	170800
Task5A	bammarkduplicates	100.5 MiB	2(normal/tumor)	2	200
Task6A	GenomonFisher GenomonMutationFilter EBFilter	41.8 MiB	30-100 (ゲノム探索領域数による) ※非DRMAA形式では逐次実行	30-100	3000-10000 非DRMAA形式では検体分はマルチプロセス
Task7A	awk	2.9 MiB	1	1	100

標準入力サンプル：バイナリーbam (4.5MiB) → 約30万行 アスキー 1_1.fastq/1_2.fastq (21MiB×2 = 42MiB)

fastqファイルの標準分割数(GenomonPipeline) = 40,000,000 = 4000万行

京向けの入力ファイル：bamファイルが500GiBと想定すると、fastqファイルは3,413,333万行
= 854 ファイルの1_1.fastq/1_2.fastqが生成 (2333GiB × 2 = 4.6TiB)

図4 Genomon2 の各タスクに対する必要メモリおよびプロセス数

引き続き、多数検体の同時解析に向けた移殖・改良設計を実施した。主に懸念される問題点は以下の二つである。一つ目はインプットファイルのサイズである。図4に現在の WGS データを基準としたインプットファイルサイズとファイル数を示す。インプットファイルの容量はWGSの場合、およそ350~450GB/検体である。100 検体を同時に解析することを仮定すると、入力データは合計 35~45TB/となる。解析の最初のステップはファイル変換またはファイル分割であり、いずれの場合も 1 検体に対して逐次実行となるため、ファイルシステムとメモリ容量の双方の観点からシステムへの要求が非常に厳しいものになる。この問題に対しては、インプットファイルの初期分割で対応予定である。次世代シーケンサーが出力するデータは WGS の場合、複数に分割されていることが多いため、これをさらに適当なサイズに分割し、それらのファイル群を 1 検体と識別するためのコントロールデータを用いて複数ノードにステージング、変換/分割処理を実施する。

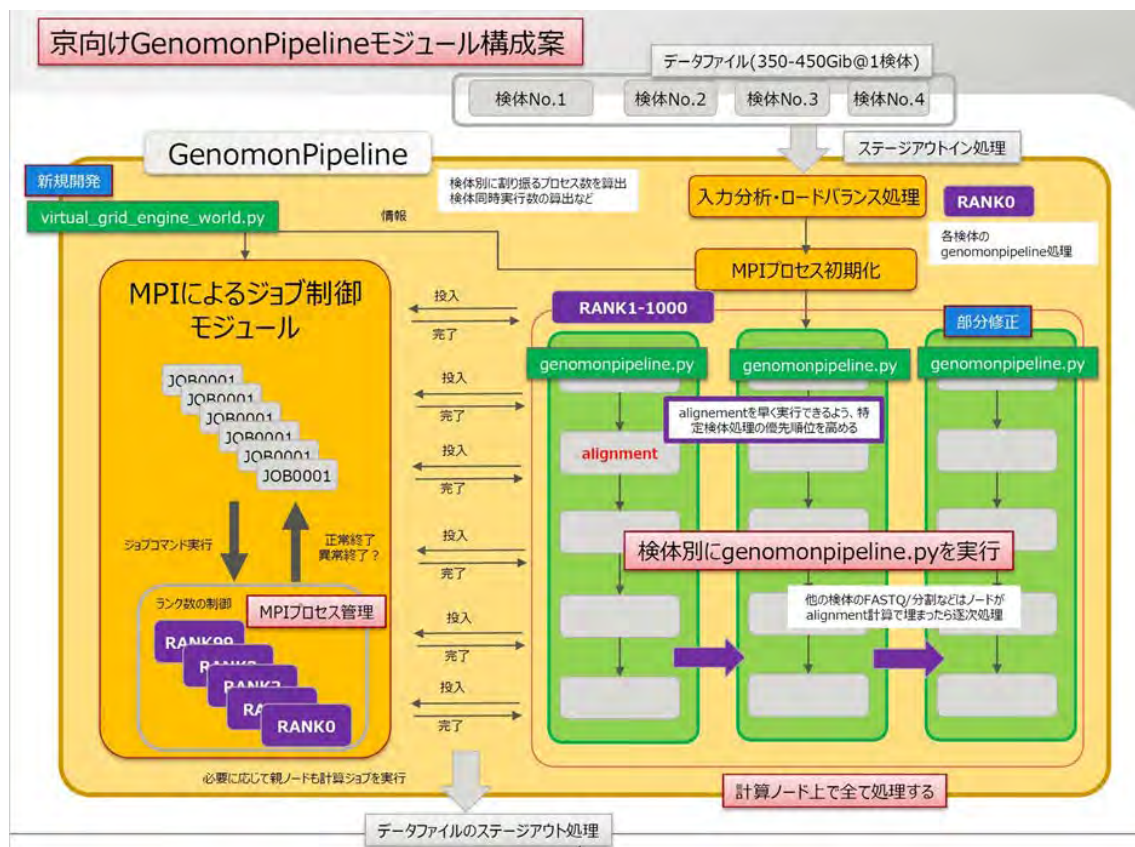


図5：バーチャル・グリッドエンジンを用いた Genomon2 の多数検体解析実施概念図

二つ目はグリッドエンジン環境を念頭に置いたワークフローの再構築である。GFK ではフローが単純であったため、並列数や処理内容を元にワークフローを3分割し、それぞれに対して専用のMPIラッパーを作成することで並列化・多数検体対応を実現した。一方、Genomon2のワークフローは複雑である。また、多数検体の解析を恒常的に実施することを考えると、ワークフローの状態管理（各検体の解析進捗状況、各タスクの計算の成否確認等）を自動的に行うシステムが必要となることは明らかである。Genomon2ではRuffus（Pythonパッケージ）を用いることで、簡易的な状態管理を実施している。また、Genomon2は現在機能拡張中であり、パイプラインの調整や新たな機能追加などが日々行われている。このような現状を踏まえ、バーチャル・グリッドエンジン（VGE）を設計を開始した。図5にVGEとそれを用いた複数検体解析のイメージを示す。Genomon2はPythonで記述されており、mpi4pyパッケージを通して各タスクをMPI並列処理する。全体はマスタスレーブ方式で処理し、Genomon2の1検体処理のジョブを複数受け付け、各検体のタスクをアレイジョブとしてマスターにサブミット、マスターがmpi4pyを通して空いているMPIランクに対してアレイジョブの内容を分割/実行させる。

2.3 ファイルシステムの問題

平成27年度の作業中に発生したファイルシステムの問題について付記しておく。Genomon2で利用する主要アラインメントソフトウェアの一つにBWA(Burrows Wheeler Aligner)がある。BWAはこれまでに京

移植が完了しているが、それを用いた性能調査の過程で「京」と Intel 環境に著しい計算時間の開きが見られた。具体的には、Intel 環境において 1 分以下で解析が終了するテストデータが、「京」での実行にはおよそ 2000 秒もかかった。

この問題は、Genomon2 の FX10 での実行時には観測されなかったため、問題の発覚が遅れてしまった。その後の調査結果から、あるひとつのバイナリファイルの読み込みに 1900 秒以上の時間が掛かっていることが原因と判明した。この現象は、ファイルシステムの I/O を行う際のクライアントキャッシュサイズが 128MB であることが原因である。一方、一般的なファイルシステム（Lustre 等）ではもっと大きなサイズが設定されており、この問題は富士通ファイルシステム FEFS 特有の問題である。コデザインにおいては、このような失態がおこなないように配慮したい。

「京」コンピュータでの計算時には、移植以外にこのような京特有の特殊仕様に起因する問題が多発しており、バイオインフォマティクス分野での利用に対する著しい障害となっている。ポスト「京」においては Linpack 性能に注力しすぎるあまりに、このような特殊仕様を採用されないことを希望することも、本準備・調査研究の結論である。

3. Genomon2 による大規模データ解析によるがんの個性と時間的・空間的多様性・起源の解明

がんの個別化医療・予防の実現のためには、個々人のがんの相違、1つのがんの内部での多様性・起源を、全ゲノムシーケンス、RNAシーケンス、メチル化シーケンス、単一細胞シーケンス、さらには血漿をソースとしたDNAシーケンスから得られる大量シーケンスデータ解析を用いて明らかにする必要がある。

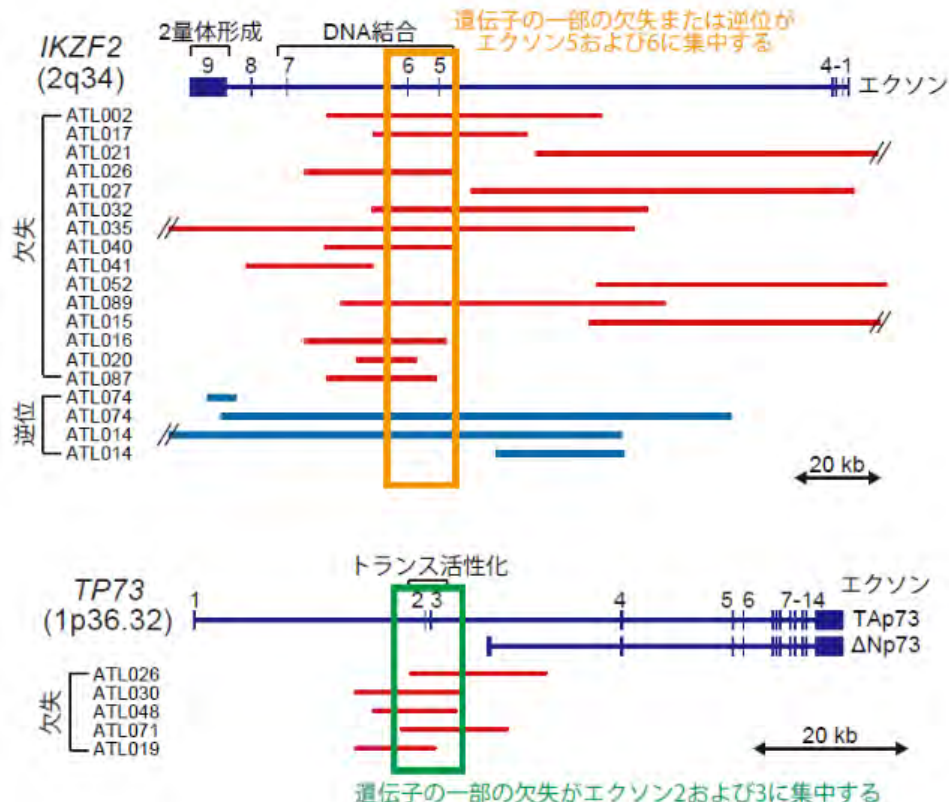


図 6：シーケンス解析の一例（ATL）

本準備・調査研究の段階では、現実のがん試料のデータに即して、「京」コンピュータ、及びポスト「京」の特段の計算能力を駆使することにより、こうした超大規模データ解析に必要な現実的な計算機ソリューションを提供することにより、ゲノムシーケンスに基づいた個別化・予防医療の実現を強力に支援するための統合計算生命科学の創設を目的に、がん試料のデータに即して、ヒトゲノム解析センタースーパーコンピュータシステム、及び「京」コンピュータ上での大規模シーケンスデータ解析環境の整備とがん研究のビッグデータ化状況について調査した。さらに、開発・整備を行った大規模シーケンス解析環境を用いて、既に得られたがん関連のビッグデータに関して、がんの個性と時間的・空間的多様性・起源の解明を目的として、下記の通り解析を実施した。全ての研究においてGenomon2の開発及びGenomon2を利用した成果である。

3.1 全ゲノムシーケンスによるがんの個性の解明

- － 成人 T 細胞白血病リンパ腫の患者各 48 ペアの全ゲノムシーケンスを行い、遺伝子変異や構造異常の同定を行った。その結果、NRXN3 や IMMP2L などの common fragile site に多発する欠失や、TP73 や IKZF2、CARD11 などの遺伝子内欠失などの新規病変を同定した。
- － 同時に RNA シーケンスデータと併せて解析することにより、異常スプライシングに関与する異常を同定することができた。特に、GATA3 や HNRNPA2B1、FAS などの遺伝子は特定の部位に繰り返しスプライシング異常を認め、ドライバー異常であると考えられた。

b) がんの時間的・空間的多様性・起源の個性の解明

- － 食道癌: 食道癌患者および同癌ハイリスク患者合わせて 32 例において、癌部、異型上皮、正常上皮より多部位サンプリングを行い、エクソーム解析ならびに標的シーケンスを実施した。その結果、異型上皮と癌における変異パターンの違いを明らかにすることが可能であった。
- － 大腸癌: 10 例以上の、大腸ポリープまたは癌を有する患者について、ポリープ(ないし癌部)および周辺健常粘膜より多部位サンプリングを行い、エクソーム解析ならびに高深度シーケンスを行った。現在、その詳細を解析中である。
- － 肝がん: 計 10 例の肝硬変・肝がんの生体肝移植症例より計 200 個以上の単一結節を採取し、エクソーム解析ならびに標的シーケンスにより詳細に解析した。その結果、肝硬変の単一結節にクローン性が認められること、一部であるが、肝がんと同様の変異を認めることが明らかとなった。

(2) データ同化生体シミュレーションによる個別化医療支援 (サブ課題 B)

調査研究・準備研究フェーズとして、サブ課題 B の実施内容について、開発計画（研究開発内容、目標・期待される成果、実施体制、必要計算資源、工程表、所要経費等）の詳細の策定を行った。また、策定した開発計画に基づき、アプリケーション開発の本格実施に向けた準備（実施体制構築の着手等）を行った。具体的には、全脳循環代謝シミュレータの開発とデータ同化イメージベーストシミュレーションによる脳動脈瘤の診断治療支援の研究開発を実施するために、3次元内部構造顕微鏡を用いてマウス脳血管の形態計測を行い、全脳循環代謝シミュレータ開発に向けた脳血管モデルの構築を開始するとともに、個別化医療支援に向けた大規模シミュレータの開発計画の詳細を策定し、アプリケーションの開発準備を行った。以下にその詳細を報告する。

1. 全脳循環シミュレータの開発（和田グループ）

（1）全脳血管モデル構築のためのマウス脳血管計測

全脳血管モデルを構築してポスト京を用いた大規模数値シミュレーションにより脳内の詳細な血流動態を再現するために、高輝度放射光（SPring-8）による CT 計測装置および 3 次元内部構造顕微鏡を用いてマウス脳血管の形状を測定した。なお、動物実験は大阪大学大学院基礎工学研究科動物実験委員会の承認（動基工 27-1-0）を得て行った。

（a）高輝度放射光 CT 計測

計測には、イソフルラン 3% で維持麻酔したマウス（C57BL/6）に、血管造影剤として硫酸バリウム（バリトゲン HD、伏見製薬）、4% アガロース溶液、マイクロフィル（MV-120）を左心室から 10 分間全身または上半身還流し、造影剤が凝固した後に頭蓋から摘出した脳（図 7）を用いた。なお、バリウム造影剤およびアガロース溶液を用いた試料は、旭川医科大学から提供いただいた。

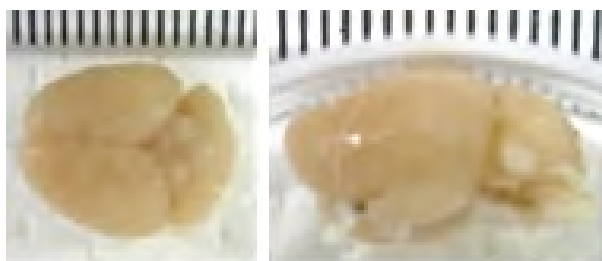
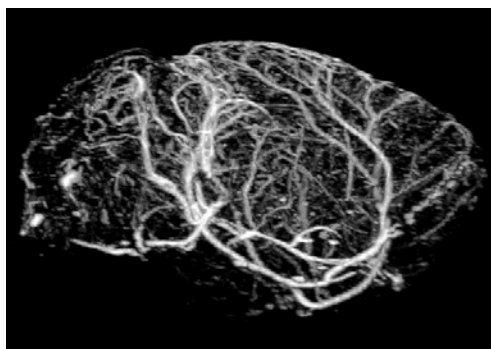


図 7： 摘出したマウス脳（上記目盛りは 1mm）

バリウム造影剤を用いた試料を用いて高輝度放射光 CT 装置で計測した例を図 8 に示す。解像度は $15\ \mu\text{m} \times 15\ \mu\text{m} \times 12\ \mu\text{m}$ である。この解像度では、脳皮質表層の細動脈レベルの血管まで抽出できているが、ところどころ切断されている血管もあり、血管抽出方法に工夫が必要であることがわかった。また、静脈系は脳摘出時に硬膜に埋没している大脳静脈が剥離され計測されていなかった。このため、頭蓋の影響を除去するためには、脱灰を行うなどの対応が必要である。また、アガロース溶液ではコントラストがつきにくく、また、マイクロフィルは粘度が他の造影剤に比べて 10 倍ほど高く、脳毛細血管レベルまで十分に還流できないなどの問題点が明らかとなった。



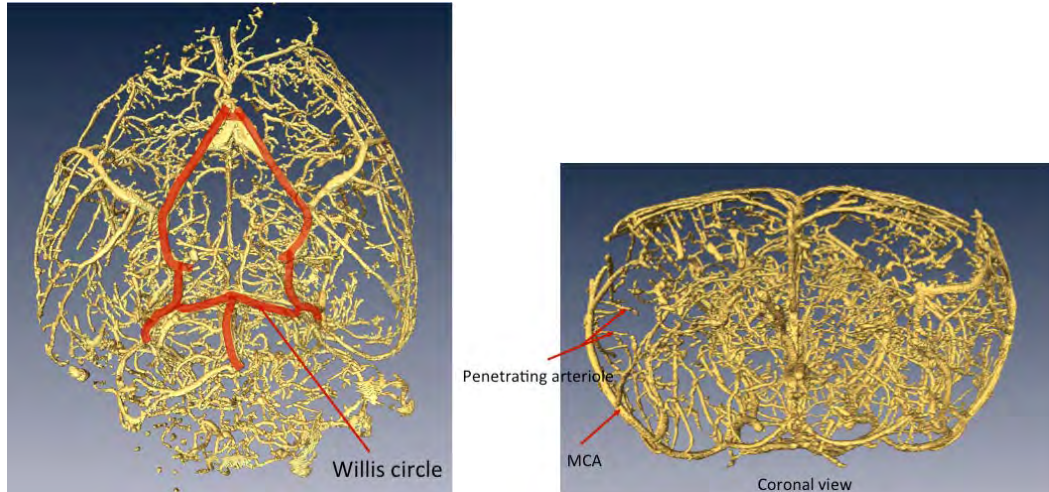


図 8 : 脳血管セグメンテーションの結果

さらに、生体組織のように弱 X 線吸収物体の中で、高感度に吸収度の違いを検知できる X 線位相差 CT による計測も試みた。造影剤にはアガロースまたはマイクロフィルを用いた。本実験では造影剤の還流が十分でなく、多数の血管を抽出できなかったが、図 9 に示すように脳実質と血管を同時に計測できることが確認できた。

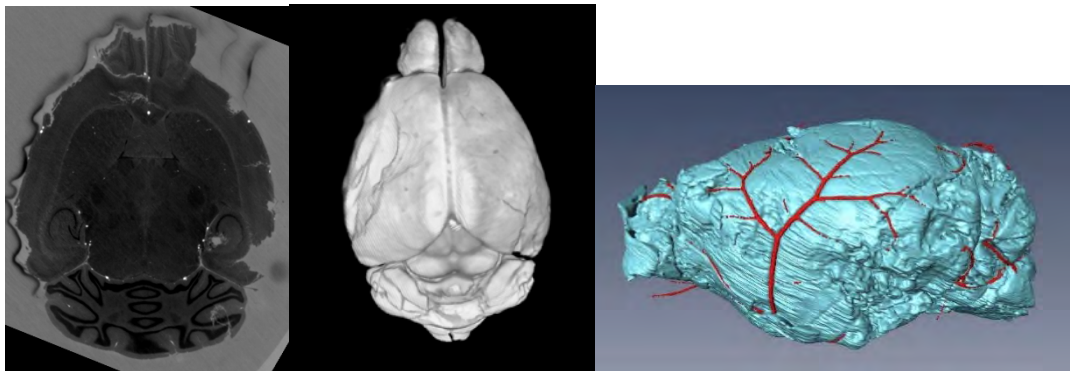
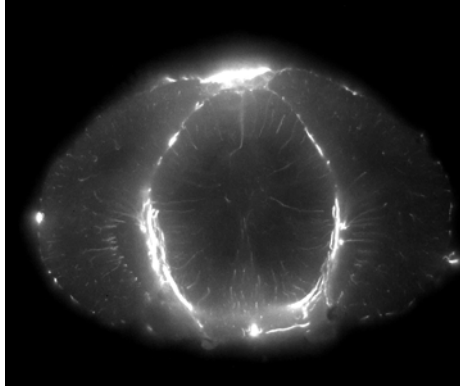


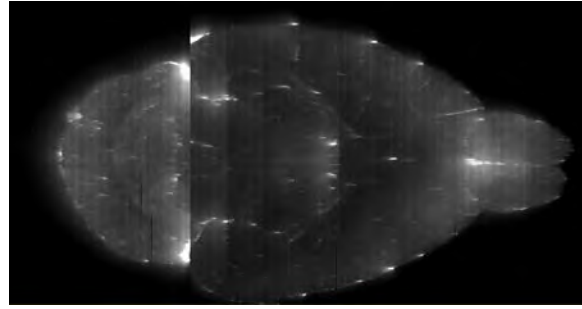
図 9 : 位相差 X 線 CT で計測された結果

(b) 3次元内部構造顕微鏡

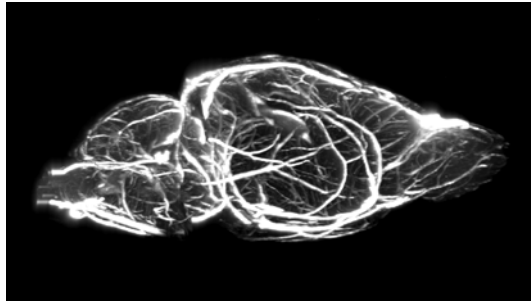
理化学研究所理化学研究所光量子工学研究領域エクストリームフォトンクス研究グループの協力を得て、3次元内部構造顕微鏡を用いて、蛍光試薬入りゼラチンを還流して固定したマウス脳の観察を行った。計測条件は、対物レンズ 0.63 倍、中間倍率 1.25 倍、ナイフ回転数 30~90rpm、切削厚 $2\mu\text{m}$ 、露光時間 200~1000 ms、フィルター $488\mu\text{m}$ DM である。今回は、試行的に実施した実験であるが、図 10 に示すように、計測範囲を絞れば、高輝度放射光 CT では観察が難しい脳皮質内部の細動脈および毛細血管網の計測が可能であることが確認できた。今後は、高輝度放射光 CT と 3次元内部構造顕微鏡の両者から得られる情報を統合して、マウス全脳血管モデルを構築していく予定である。



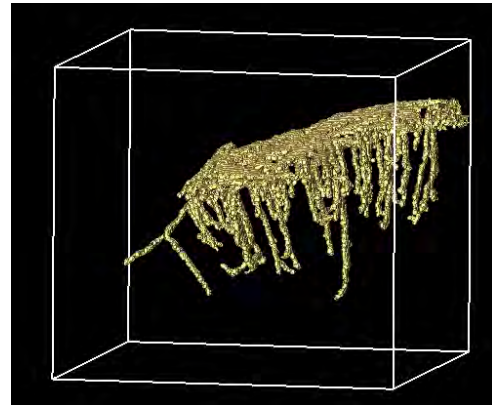
(a) 断面画像 (Coronal)



(b) 断面画像 (Sagittal)



(c) 3次元画像 (全体)



(d) 3次元画像 (大脳皮質内部)

図 10 : 3次元内部構造顕微鏡で得られたマウス脳血管形態

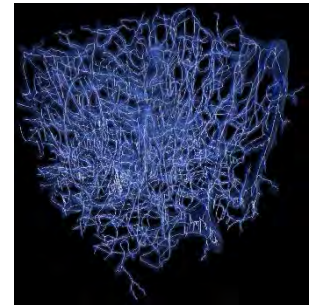
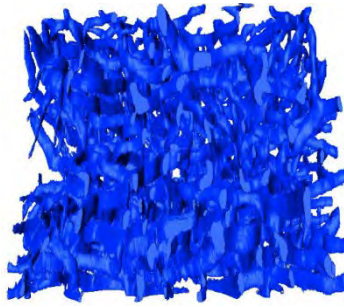
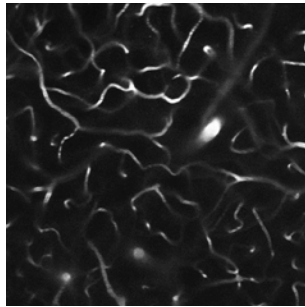
(2) 脳微小循環血流シミュレーション手法の開発

血管新生の数値モデルを導入しボクセル格子上で定式化および離散化することにより、スーパーコンピュータ「京」を用いた大規模計算による脳微小循環網の形態モデル構築を達成した。この際、医用計測画像から得られる情報を融合させることにより、計測データを反映しつつ不完全な計測データを数値モデルで補完した脳微小循環網形態が得られることを示した。また、 1000^3 ボクセル程度の条件においては、容易に形態構築が可能であることを示した。本手法により得られた形態データはボクセル格子で定義されているため、直接、血流シミュレーションの形態情報として用いることが可能であり、脳微小循環シミュレーションの今後の展開を示すことができた。

(a) 計算モデル

二光子顕微鏡により得られたマウス大脳皮質内毛細血管網のスライス画像 (1024^2 ピクセル) および画像処理ソフト Amira (Visage Imaging, Berlin, Germany) を用いて抽出された血管形状とその中心線を図 11 に示す。図 11 (b) の抽出された形状を見ると、二光子顕微鏡の焦点特性のため深さ方向に形状が伸長している。毛細血管においては粘性流れが支配的であり血管形状が流動抵抗に大きく影響するため、このままでは血流解析に直接用いることができない。一方、抽出された中心線を見ると (図 11 (c))、細線化処理の特性より、血管形状の伸長に依存せず適切に抽出が行われていることが確認できた。そこで本研究では、この中心線を構成する点群の情報を活用した形態モデルの構築を行った。また、ボクセ

ル格子を用いた血管新生モデルによるアプローチを導入することで、「京」のシステムにおいて、形態モデル構築と血流解析のシームレスな解析を試みる。



(a) 二光子顕微鏡により取得した
スライス画像

(b) 画像処理により抽出された
血管形状

(c) 細線化により抽出された
中心線（中心点群）

図 1 1 : マウス大脳皮質内の毛細血管網の画像および画像処理ソフト Amira による再構築結果

血管新生の巨視的数理モデル[1]では、空間的に配置された低酸素腫瘍細胞から血管新生因子を空間的に拡散させることで既存血管に到達させ、内皮細胞との反応・確率モデルを用いて血管新生を引き起こす。本モデルでは、相変化問題として現象を取り扱い Phase-field 法を用い定式化を行っているが、4 階の空間微係数を含むため演算に必要なステンシルが広がり並列化の弊害となる。これに対し本研究では、Phase-field 法を必要としない新たな定式化により、大規模並列化に適したアプローチを提案した。以下にその概要を示す。

・血管新生因子の濃度 $c(x, t)$ の輸送

$$\partial_t c = DV^2 c - \alpha c f + S \quad (1)$$

左辺第一項は時間変化項、右辺第一項は拡散項、右辺第二項は血管相における血管新生因子の消費項、 S は低酸素細胞からの発生項を表し、 D は拡散係数、 α は消費強さを決定する係数である。発生項に関して、図 1 2 に示す抽出した中心点群を低酸素細胞と見なし血管新生因子を放出させる。

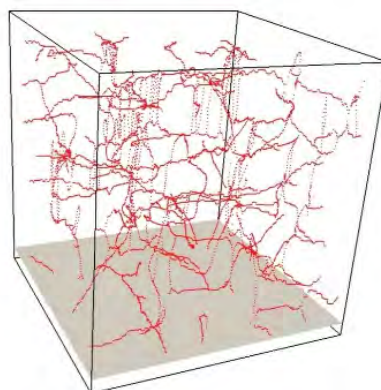


図 1 2 : 抽出された血管中心点群（赤点）

・血管を記述する相関数 $f(x, t)$ の輸送

$$\partial_t f + \mathbf{v} \cdot \nabla f = 0 \quad (2)$$

式(2)は移流方程式であり、移流速度（あるいは血管成長速度） $\mathbf{v}$ により相関数が輸送されていく。ここで、血管新生・成長メカニズムを簡易的に考慮し、 $\mathbf{v}$ に関して以下のモデル化を行った。

$$\mathbf{v} = \beta L(|\nabla c|) \frac{\nabla c}{|\nabla c|} \quad (3)$$

$$L(|\nabla c|) = \begin{cases} 0, & \text{for } |\nabla c| \leq |\nabla c|_{th} \\ 1 - \frac{|\nabla c|_{th}}{|\nabla c|}, & \text{otherwise.} \end{cases} \quad (4)$$

$\square$ は成長速度を決定づけるパラメータ、 $L(|\nabla c|)$ は新生因子濃度 c の空間勾配 $|\nabla c|$ の大きさに依存した関数であり、 $|\nabla c|_{th}$ は血管新生が起きる閾値である。式(4)より濃度勾配の絶対値が閾値を上回った際に血管新生が起こり、式(3)よりその成長速度は濃度勾配およびその向きに依存するモデリングになっていることが分かる。式(1)の時間離散化に一次精度オイラー陰解法、空間離散化に二次精度中心差分法を適用した。また、式(2)では境界領域の数値散逸を抑えるために界面捕獲手法に一種である MTHINC 法[2]を適用した。

(b) 並列計算の方法と効果（性能）

直交領域分割による MPI+OpenMP のハイブリッド並列を行った。本離散化において、式(1)では近隣一つ、式(2)には近隣二つまでのステンシルが必要となる。そこで、各担当ノードに自身の領域に加え境界領域を設定し変数を保持し、必要に応じて隣接ノード間で各変数を非同期通信した。また、濃度輸送には陰解法を用いてるためポアソン方程式を解く必要があり、本モデルでは対角スケーリング付き Bi-CGSTAB 法を用いた。この際、各担当ノード内でベクトル・行列積演算を行い、必要に応じて隣接ノード間で非同期通信、また全領域での和が必要とされるスカラー値を集団通信により処理した。これらの結果、910*910*700 ボクセルにおいて定常状態まで（およそ 10 万ステップ）の計算が 1000 ノードを用いて 1 時間程度で終了した。これは十分現実的な解析時間である。詳細な単体性能および並列性能の調査は今後検証していく予定である。

(c) 結果

解析領域の一部における本提案モデルを用いたシミュレーション結果を図 1 3 に示す。計算ステップを増すごとに、中心点群を覆いながら血管が新生・成長した。また、画像処理による抽出形状（図 1 1 (b)）と比べると、形状が深さ方向に伸長していないことが確認できた。

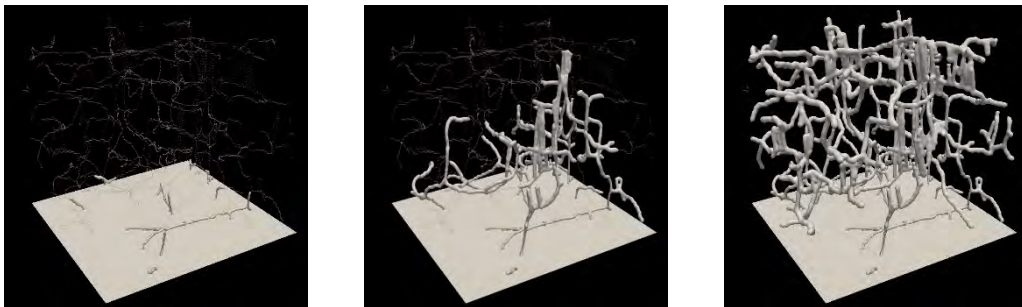
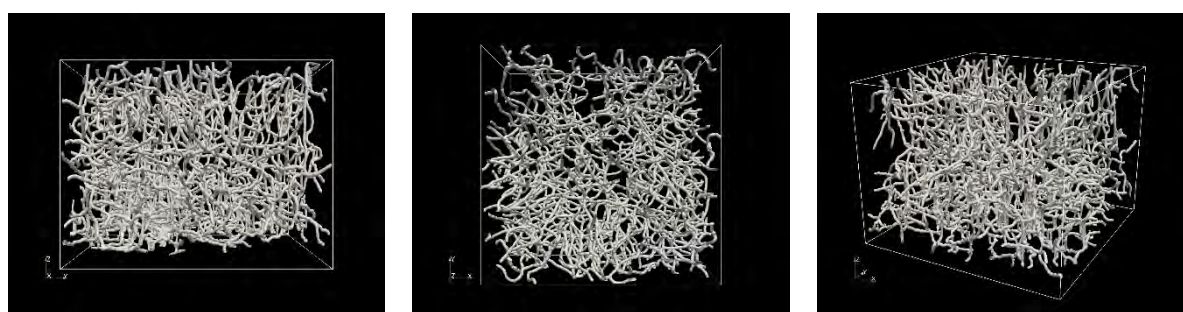


図 1 3 : 提案モデルによる血管新生・成長の様子

次に、 $910 \times 910 \times 700$ ボクセル全体での成長後の結果を図 1 4 に示す。「京」を利用することにより、わずか数マイクロメートル径の毛細血管を有限格子で捉え、これら血管群からなる複雑な網構造を再現することができた。しかしながら、深さ方向の伸長を避けるために、血管表面の情報を利用しておらず、血管径がすべて均一となってしまった。これらは、観察精度が保証されている平面内での血管径情報をモデルパラメータに反映させることで改善が見込まれ、今後、血管径のバラつきを考慮した形態モデルの構築が可能になると考えられる。



yz 面視点からの図

xz 面視点からの図

斜めからの鳥瞰図

図 1 4 : $910 \times 910 \times 700$ ボクセルにおける毛細血管網の再構築結果

(d) まとめ

データ同化生体シミュレーションによる個別化医療支援の実現を目指し、本研究では、脳微小循環を対象とし、医用計測画像に基づいた大規模計算による形態モデル構築手法の提案を行った。血管新生の数理モデルを導入しボクセル格子上で定式化および離散化することにより、「京」コンピュータを用いて、 1000^3 ボクセル程度の条件で容易に形態構築が可能であることを示した。本手法により得られた形態データはボクセル格子で定義されているため、直接、血流シミュレーションの形態情報として用いることが可能である。今後、「京」を用いた脳微小循環血流解析において、形態モデル構築と血流解析のシームレスな解析を行っていく予定である。

参考文献

- [1] R. D. M. Travasso, et al., PLoS One, 6, e19989 (2011).
- [2] S. Ii, et al., J. Comput. Phys., 231, 2328–2358 (2012).

2. 医用画像データを基にした階層統合生体力学シミュレーション（高木グループ）

脳血管 Willis 動脈輪の様に多数の分岐を伴う複雑な血管網に対して 3 次元大規模計算と 0 次元–1 次元血管網シミュレータを連成させるシミュレーション手法の開発を進め、実際に脳血管系の画像データを用いたシミュレーションによりその有効性を検証した。また、血栓シミュレータの開発と関連して、血小板の画像データを基にしてシミュレーションを実施する手法の開発を進めた。さらに、脳血管における血液脳関門 (Brain Blood Barrier (BBB)) に対して薬剤を透過させる技術として注目されている超音波照射手法に関連して、血流中のマイクロバブル内包型ベシクルに対して超音波を照射し、マイクロバブルの振る舞いを解析するためのシミュレーションツールの開発を進めた。

神経系-筋骨格系の連成問題に関しては、大阪大学の野村泰伸教授と連携して、これまで「京」向けに開発を進めてきた神経細胞・筋繊維レベルからの筋骨格系の階層統合シミュレーション手法に改良を加

え、下半身の筋骨格系画像データを取り込んだ大規模シミュレーションに向けた準備を進めた。

3. 微小循環遊走細胞解析プラットフォームの開発準備（松田グループ）

微小循環遊走細胞解析プラットフォームの開発とデータ同化イメージベーストシミュレーションとの連携に向けて、平成27年度は、生体内部の組織を生きた状態で直接観察できる二光子励起顕微鏡を用いて、マウス微小血管中での好中球等の白血球細胞を蛍光標識して撮影した経時観察動画像を、健常時と刺激剤により炎症を誘発した状態のそれぞれについて取得した。さらに、この二光子励起顕微鏡画像の解析により、血管中で特徴的な動態を示す血球細胞を検出する手法について検討を行った。

具体的には、二光子励起顕微鏡により生体内を経時的に毎秒30フレームで画像を撮影できるため、時間的に連続する画像フレーム間で類似の輝度を持つ画素どうしを対応付けることで、画素の速度ベクトルを算出するオプティカルフローを用いて、血管中の血球細胞の移動速度の計測を試みた。一般に、経時観察画像から細胞を認識して動きを追跡するには、画像中で細胞の占める領域を認識するセグメンテーションや連続する画像フレームから細胞の動きを求めるトラッキングの処理が必要になるが、非常に複雑な生体内の血管画像から細胞の領域を正確に認識しその動きを追跡するのは容易ではない。しかし、本研究では、血栓や動脈硬化の原因となる血管内皮組織の変化に着目しているため、個々の血球細胞ではなく血管内皮組織の位置ごとの血球細胞との相互作用が血球細胞の動態に与える影響について明らかにすることを目指しており、個々の血球細胞の詳細な追跡は必要ないため、オプティカルフローに基づく手法による移動速度計測を考えた。

このための準備として、オプティカルフローを用いて画像内の各画素の速度ベクトルを求める手法を「京」上で実装した。本実装では、コンピューター・ビジョン・ライブラリ OpenCV 中の Gunnar Farneback のアルゴリズムに基づく密なオプティカルフローを求める手法を用いた。さらに、得られたオプティカルフローの速度ベクトルから、血球細胞の移動速度に対応する速度ベクトルの抽出を行った。「京」上での並列処理は、多数の画像フレームからオプティカルフローを求める処理を、MPIによりデータ並列による並列化として実現した。

マウス生体の血管を二光子励起顕微鏡で観察した動画像に対して本手法を適用した結果、血球細胞の移動速度は均一でなく血管中の位置ごとに偏っていることが確認できた。特に炎症発生時には血球細胞が血管壁の特定の部位に張り付いて留まるという現象があるとされており、本研究の解析結果でも血管中の位置ごとの細胞移動速度の偏りが健常時に比べて顕著であることが計測できた。

また、炎症による血管内皮組織の状態変化を分子レベルでも解析するため、細胞内の遺伝子発現と細胞動態の関連解析について検討を行った。具体的には、血管内の個々の血球細胞の遺伝子発現量を RNA シーケンスにより取得して、複数の血球細胞間で横断的に比較解析する手法の開発を目指している。平成27年度は、遺伝子発現量の計測時に生じる細胞ごとのばらつきを補正し、細胞間で定量的に特定の遺伝子の発現量を比較可能とするための発現量の正規化手法や発現変動遺伝子探索等の解析手法について検討した。

今後は、血栓や動脈硬化と関連があるとされている炎症の血管内皮組織に及ぼす影響を細胞イメージングと分子レベルの両方面から解析する手法について研究開発を進める予定である。

4. 大規模計算ニューロメカニクス研究基盤プラットフォームの開発準備（野村グループ）

脳神経系は、身体機械力学系としての筋骨格系の動特性を巧みに利用することで、柔軟性を保ちながらも安定性の高い身体運動を達成している。したがって、脳神経と身体機械力学系、および外界が相互作用の結果として実現されるヒト運動制御における上位運動脳の役割、あるいは脳疾患に起因する運動障害の機序を定量的に明らかにするためには、脊髓神経と身体機械力学系の精緻な数値モデルの構築が不可欠である。この課題では、「京」で開発してきた神経系-筋骨格系統合シミュレータを発展させ、上位運動脳（大脳皮質・基底核・小脳・視床・脳幹）の制御対象としての超大規模脊髓神経・筋骨格系モデルを開発し、ヒト脳による運動制御（特に立位姿勢および二足歩行運動の安定化制御）メカニズムおよび神経疾患に起因する運動障害の発現メカニズムの解明に資する大規模計算ニューロメカニクス研究の基盤プラットフォームの開発を行っている。このプラットフォームは、パーキンソン病等の中枢神経疾患に起因する運動障害の定量診断および定量的薬効評価を可能にし、先端医療の推進に貢献する。具体的には、東京大学高木グループと連携し、戦略プログラム（分野1課題3）で開発されたX線CTおよびMRIを用いた静的身体幾何学モデルに対して、3次元身体運動計測データと筋活動計測データを付加し、データ同化技術を適用することで、運動に関わる多数の筋群に対する上位運動脳による神経制御信号の推定（逆動力学神経ネットワーク同定）を可能にする精緻な脊髓神経・筋骨格系モデルを構築すること、さらに、パーキンソン病患者の立位姿勢および歩行に対して開発した数値モデルを当てはめ、運動計測に基づく神経制御信号の逆推定を行うことを目指している。平成27年度は、戦略プログラムで開発してきた筋モデルの精緻とソフトウェアの整備を実施した。特に、足関節筋のマルチスケールモデルの開発に焦点を絞り、現在、筋収縮モデルとして、状態遷移理論に基づくアクチン分子とミオシン分子のクロスブリッジ形成を考慮したサルコメア力学モデルの導入を行い、医用画像と文献から得られた材料定数に基づいて、サルコメアから筋群内の筋線維走行までを精緻に再現するモデルを構築しているところである。

5. 実形状口腔流体音響シミュレータ開発に向けた単純形状モデルによる検討（玉川グループ）

歯科治療において、口腔の形態異常により発生する発音障害は一つの大きな問題となっており、治療後に発生する障害を治療前にあらかじめ予測・予防することが求められている。発音の中でも摩擦音は、口腔内に小さな狭めを形成する事により発生するジェット流から発生するため、歯科治療による口腔形態の変化に大きな影響を受ける。しかし、ジェット流による音源発生と口腔内での音の共鳴という複雑な物理現象が組み合わさって音が形成されるため、治療前後の音の変化を歯科医師の経験から予測することは困難である。そこで、本研究では歯茎摩擦音の口腔単純モデルを用いて数値流体シミュレーションを行うことにより、発音時の口腔内に発生するジェット流の形態に関する予測を試みた。計算は、京の利用により可能となる大規模計算格子(3億6千万格子)を用いた。

本研究では三次元非圧縮性 Navier-Stokes 方程式を基礎方程式とし、乱流モデルとして Dynamic Smagorinsky モデルを用いたラージエディシミュレーション(LES)を行った。計算は、時間・空間ともに2次精度の有限要素法ソフトウェアFrontFlow/blue ver. 8.1を用いた。口腔単純モデルは咽頭、狭窄流路、上下前歯、口唇、遠方場から成り、6面体一次要素を用いている。格子外観を図15に示す。計算格子数は3億6千万格子である。

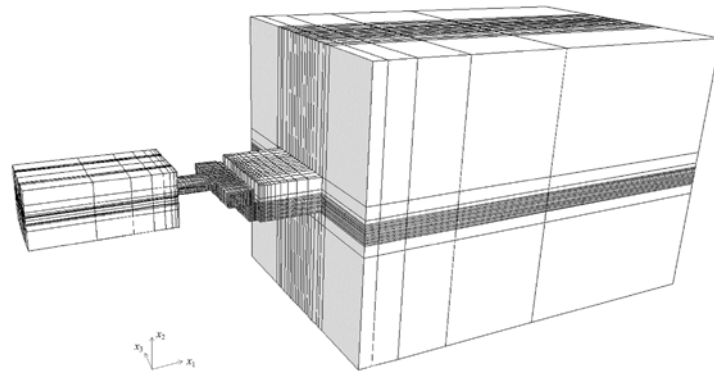


図 15 : 格子外観

口腔モデル内の正中矢状断面における流れ場を図 16 に示す。左端の咽頭部から舌と上顎の狭窄流路で流速が高まり、狭窄流路から出たジェット流が上下前歯の隙間でさらに乱れる様子がわかる。また、図 17 にジェット流の拡大図と計算格子を示す。ジェット主流から剥離する渦や壁面に当たった流れが形成する渦を解像できている様子がわかる。本計算により、発音における気流の動態を予測するためには大規模な数値流体シミュレーションを行う必要があることが示唆された。今後は患者個別の口腔形状を想定した複数のモデルに対してシミュレーションを行い、口腔形状の変化が流れと音にどのように影響を与えるのかを調べ、医療用画像の統計的機械学習による形態と疾患との関連性の発見につなげていく予定である。

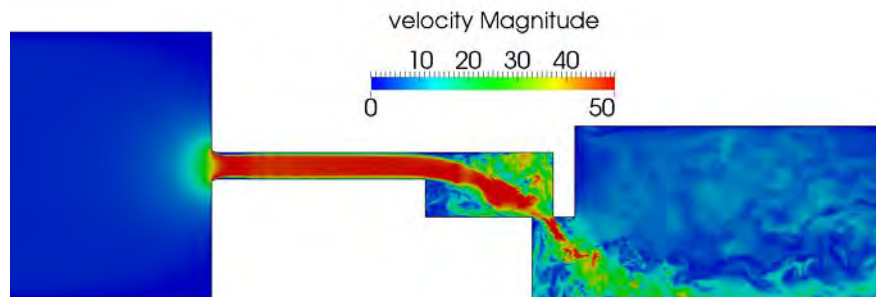


図 16 : モデル全体の流速分布

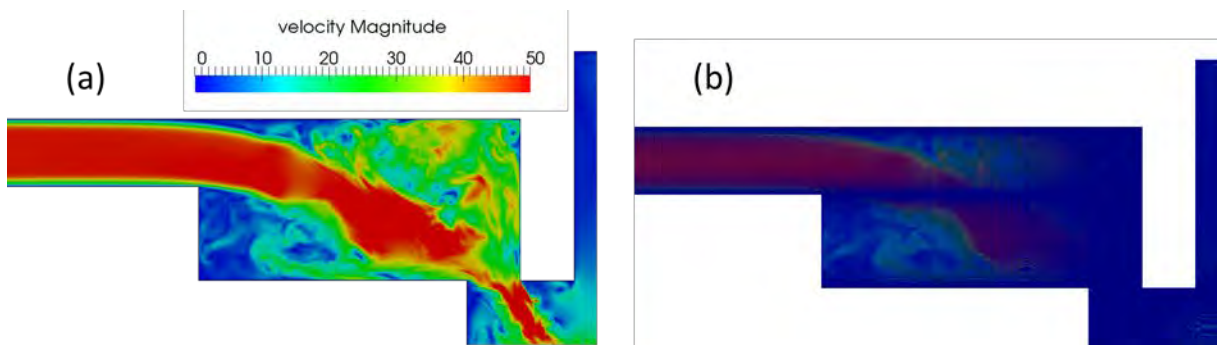


図 17 : ジェット流の拡大図(a)と計算格子(b)

(3) 心臓シミュレーションと分子シミュレーションの融合による基礎医学と臨床医学の架橋 (サブ課題 C)

調査研究・準備研究フェーズとして、サブ課題 C の実施内容について、開発計画 (研究開発内容、目標・期待される成果、実施体制、必要計算資源、工程表、所要経費等) の詳細の策定を行った。また、策定した開発計画に基づき、アプリケーション開発の本格実施に向けた準備 (実施体制構築の着手等) を行った。以下では、本サブ課題で掲げる二つの目標に大別してその内容を具体的に説明する。なお基盤技術である UT-Heart シミュレータ本体の継続的な高度化は国際競争力維持のために必須であり、それに関わる進捗は発表論文リストに譲る。

目標 1 は、心臓シミュレータ UT-Heart と分子シミュレータ CafeMol を融合させることにより、ミクロ・マクロ間の相互作用により病態が進行する心不全の解明と最適治療を可能とする世界でも例を見ないマルチスケール心臓シミュレーションを実現することである。本年度は先ずミクロ側であるサルコメアモデルの対象分子検討、並びに UT-Heart と CafeMol の連成理論の検討を行った。対象分子としては、ポスト「京」が未だ存在しない 2 年後に於ける中間目標として、最小限のアクトミオシン系を考えることとした。二つの重鎖と四つの軽鎖からなる六量体であるミオシン分子については、最小機能単位であり、疾患に関連した変異が多く報告されているミオシンサブフラグメント 1 (S1) と二つの軽鎖のみをアミノ酸粒度での粗視化分子モデルで表し、一方、計算負荷軽減の観点から残りの重鎖領域 (サブフラグメント 2 [S2] とライトメロミオシン [LMM]) は分子生理実験の結果を再現したばねで近似する方針とした。これによって他のサルコメア構造と連結し、細胞、組織レベルの力学との連成を達成する。また Protein Data Bank や文献等から、ヒトや他の動物種に関するアクチン、ミオシンの X 線結晶構造解析データの調査を行った。その結果、入手出来るヒトの構造データやクロスブリッジの状態数、更には強結合状態における界面構造の情報は限られているが、ヒトと動物種との遺伝子配列相同性は高く、動物の構造データを鋳型にしたモデル化が可能であろうことが分かった。またアクトミオシン系ではヒト心臓の 1 心拍に相当する約 1 秒間に何回の加水分解サイクルが繰り返されるかの化学反応速度についても文献調査した。調査は今後共継続するが、以上の第 1 回目調査並びに検討を踏まえ、常識的な「京」のノード時間積の範囲内でマクロ心筋モデルとの連成シミュレーションが可能なミクロモデルの自由度を先ず見積もる試算を次年度から開始することとした。粗視化分子モデルの概念を図 18 に示す。アクチンフィラメントは赤で、ミオシンのサブフラグメント 1 (S1) は緑で、二つの軽鎖 (制御、必須) については各々青とオレンジで色付けした。

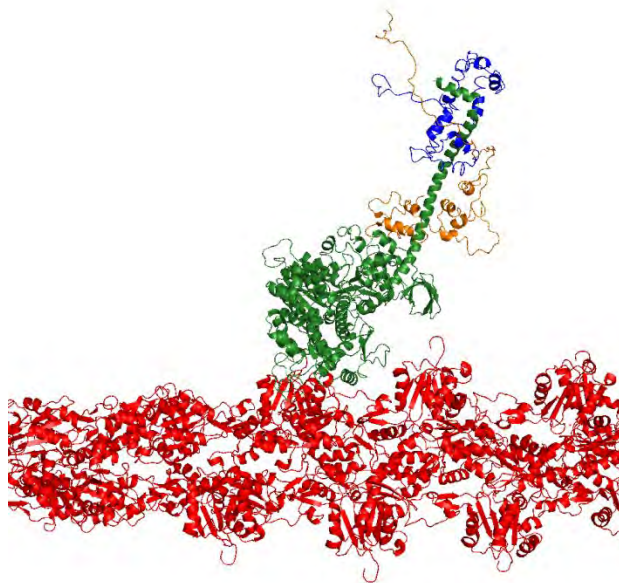


図 18 : アクトミオシン系の分子シミュレーションモデルの一例

マクロ・ミクロ連成解析法については、UT-Heart チームと京都大学高田教授、金田研究員との合同ミーティングを開催し情報交換、議論を行った。有限要素法に基づくマクロ側が陰解法、分子シミュレーションに基づくミクロ側が陽解法であり、また時間刻みについてはマクロ側がミリ秒、ミクロ側がピコ秒のオーダーであることを考慮し、直接的な連成だけでなく自由エネルギー地形を介した連成など多角的な検討を継続することとした。

マクロの血行動態と分子レベルの生理学との連成したシミュレーションの医療への応用を展開するために、医学系のグループと共に当初策定した開発計画を見直し、アプリケーション開発の本格実施に向けた準備を行った。即ち、成人の心不全に対する心臓再同期療法（CRT）を中心とした介入が予後に与える影響については自治医科大学さいたま医療センター循環器科と、小児の先天性心疾患に対する手術的な介入が予後に与える影響については岡山大学心臓外科とそれぞれ打ち合わせを行った。共にシミュレーションに必要なデータがどこまで臨床現場で取得可能であるかについて個人情報保護との関連も含めて検討し、自治医科大学、岡山大学の承認を得た。併せて「京」利用に当たり、「ヒト由来情報の取り扱いに関するチェックリスト」を作成し AICS に提出して承認を得た。更に、従来の UT-Heart のサルコメアモデルを用いたシミュレーション結果の検証を進めた。

目標 2 は、重点課題 1 との協同により、心臓シミュレータ UT-Heart と心筋細胞イオンチャネルの分子シミュレーションを融合することにより、創薬における候補物質の分子構造と遺伝子情報のみから不整脈のリスク評価を可能とするシステムを開発することである。具体的には、既に開発済の図 19 に示される UT-Heart と細胞パッチクランプ実験と組み合わせた「ハイブリッド心毒性スクリーニングシステム」のパッチクランプ実験の部分を分子シミュレーションに置き換えることになるが、現在のシステムでの結果（図 20）：UT-Heart チーム・エーザイ・東京医科歯科大, Science Advances, May01, 2015）を分

析すると、例えば図21に示されるように、Verapamilはdoseを増しても心電図のQT間隔は延長するものの不整脈は発生せず、一方、CisaprideはQT間隔の延長と共に、常用量の30倍でTdP（致死性不整脈に繋がる頻脈）が発生する。これは従来のガイドラインで危険とされるQT間隔延長が不整脈リスクの指標として必ずしも有効でないことを示唆しており、一つの薬剤により複数のイオン電流が重畳抑制されたことに起因する結果であると考えられる。では、どのような複数のイオン電流の重畳抑制がどのように不整脈発生に影響を及ぼすのか、その全体像を明らかにしておくことは、次年度以降、分子シミュレーションでモデル化するイオンチャネルの選択にも影響してくる。

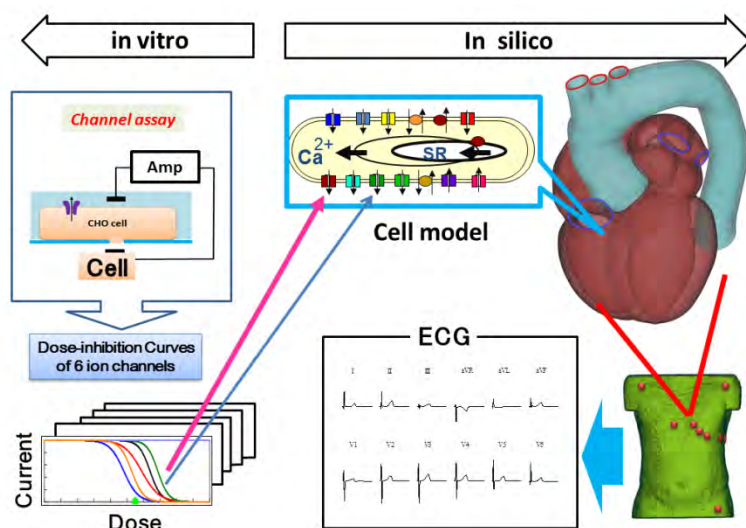


図19 開発済のハイブリッド心毒性スクリーニングシステム

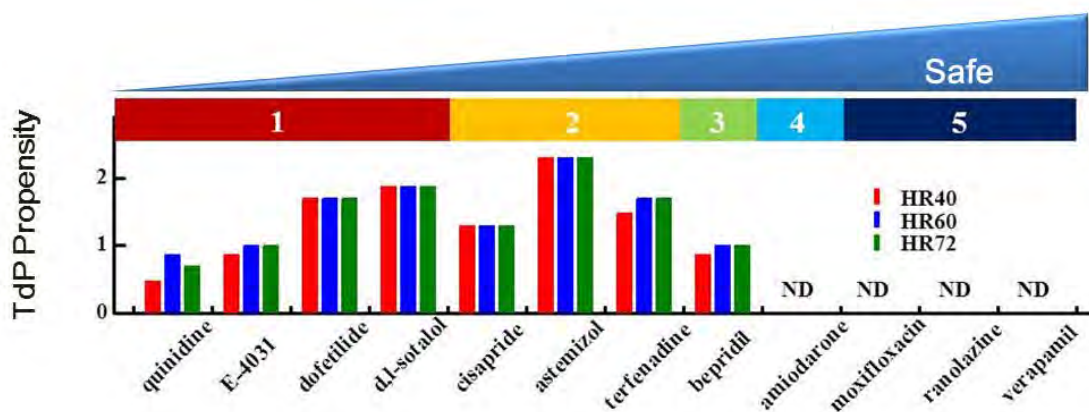


図20 ハイブリッド心毒性スクリーニングシステムによる12薬剤の解析結果 (Okada et al., Science Advances, May01, 2015)

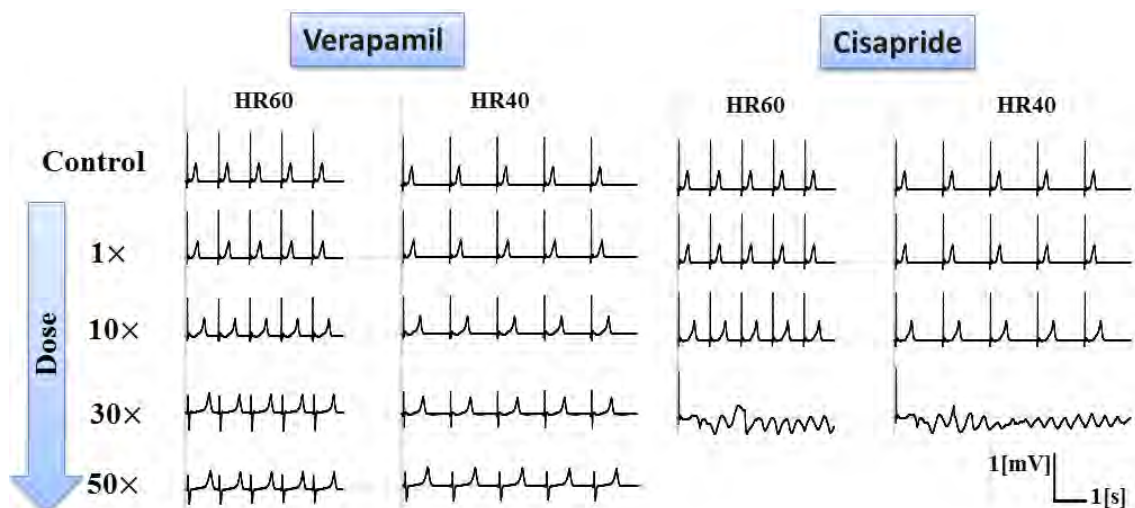


図 2 1 : Verapamil と Cisapride に対する第Ⅱ誘導心電図 (HR:心拍数)

そこで本年度は事前調査として各種イオン電流の重畳抑制が不整脈発生に及ぼす全体的な傾向を把握する計算を「京」を用いて実施した。即ち不整脈に関係の深いことが知られる、遅延整流性カリウム電流 I_{Kr} 、 I_{Ks} 、ナトリウム電流 I_{Na} 、 I_{NaL} 、カルシウム電流 I_{CaL} の 5 電流を種々の抑制率で組み合わせて UT-Heart による不整脈発生の有無を調べた。また、その際これまで提案されている各種の心電図指標 (図 2 2) も出力した。

図 2 3 に I_{Kr} と I_{CaL} を抑制した結果の一例を示す。黒色領域は不整脈発生領域、他の色は心電図指標の一つである $T_{peak}-T_{end}$ の変化量 (ms) を示す。これより、 I_{Kr} の抑制は不整脈をもたらすが、加えて I_{CaL} の抑制が 20% を超えると I_{Kr} の効果を緩和することが分かる。更に、 $T_{peak}-T_{end}$ 値は不整脈発生領域に向かって一旦減少し、その後増加に転じる傾向を示していること、並びに、不整脈発生直前での $T_{peak}-T_{end}$ の限界値は一定ではない、ことから不整脈発生を予測する有効な心電図指標とはなり得ないことが分かる。本検討については、次年度上期に計算を完了し結論を導く予定である。本検討結果は、今後の分子シミュレーションのモデル化に資するものであるが、同時に製薬企業の従来の薬理実験に基づく心毒性スクリーニングにおいて活用可能な新たな参考データにもなり得るものと考えられる。なお、本検討はポスト「京」重点課題推進 WG 委員からの「従来の薬理実験との共存を目指すのが良い」とのコメントに合致するものであることを付言したい。

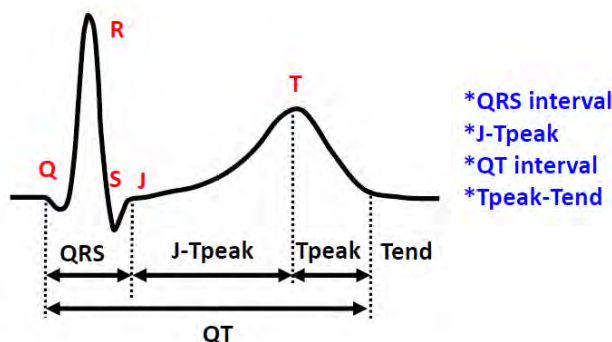


図 2 2 : これまでに提案されている各種の心電図指標

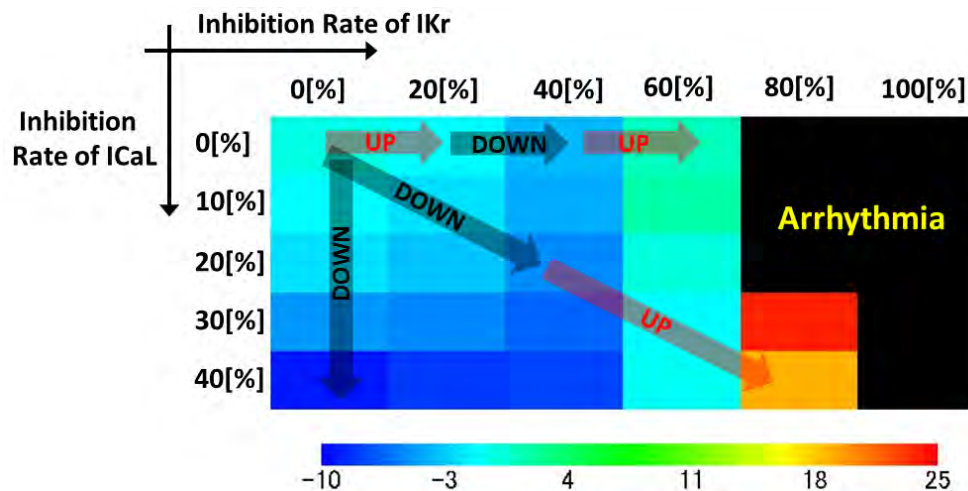


図 2 3： IKr と ICaL 電流の各種抑制率の組み合わせに対する不整脈発生の有無と Tpeak-Tend 変化量 (ms)

(4) プロジェクトの総合的推進

プロジェクトの総合的推進のため、テレビ会議、電話会議、及びメール会議による全体会議・運営委員会を開催し、プロジェクト全体の進捗状況を確認しつつ、平成 28 年度からの本格実施フェーズに向けて、今後の研究内容および連携・協力体制について調整し、実施計画書案を作成した。そして、外部有識者による諮問委員会を設置し、メール審議により、本格実施フェーズの実施計画について審議し、実施計画書を作成した。また、ポスト「京」重点課題推進 WG、重点課題連携推進会議準備会議、重点課題アプリケーション連携推進委員会、人材育成タスクフォース、サーバ等に関するワーキンググループ等、全重点課題に関する会議において連携・調整を行った。また、生物物理学会、分子生物学会等において、本重点課題の内容について発表を行った。戦略分野 1 と連携して、ニュースレターで本重点課題について内容を紹介した。ホームページの運営をした。

4-3. 実施体制

業 務 項 目	実 施 場 所	担 当 責 任 者
(1) 大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明 (サブ課題 A)	東京都港区白金台 4-6-1 国立大学 法人東京大学 医科学研究所ヒト ゲノム解析センター	ヒトゲノム解析センター長 宮野 悟
	京都府京都市左京区吉田近衛町 国立大学法人京都大学 大学院医 学研究科	大学院医学研究科 教授 小川 誠司
	大阪府豊中市待兼山町 1-3 国立大学法人大阪大学 大学院基	大学院基礎工学研究科 教授 和田 成生

(2) データ同化生体シミュレーションによる個別化医療支援 (サブ課題B)	礎工学研究科 東京都文京区本郷 7-3-1 国立大学法人東京大学 大学院工学系研究科	大学院工学系研究科 教授 高木 周
	大阪府吹田市山田丘1-1 国立大学法人大阪大学 大学院情報科学研究科	大学院情報科学研究科 教授 松田 秀雄
	大阪府豊中市待兼山町1-3 国立大学法人大阪大学 大学院基礎工学研究科	大学院基礎工学研究科 教授 野村 泰伸
	大阪府吹田市山田丘1-8 国立大学法人大阪大学 歯学部附属病院	歯学部附属病院 准教授 玉川 裕夫
(3) 心臓シミュレーションと分子シミュレーションの融合による基礎医学と臨床医学の架橋 (サブ課題C)	東京都世田谷区野沢3-25-8 株式会社UT-Heart研究所	代表取締役会長 久田俊明
	栃木県下野市薬師寺3311-1 学校法人自治医科大学	学長 永井 良三
(4) プロジェクトの総合的推進	岡山県岡山市北区鹿田町2-5-1 国立大学法人岡山大学 大学院医歯薬学総合研究科	大学院医歯薬学総合研究科 教授 佐野 俊二
	国立大学法人東京大学 医科学研究所ヒトゲノム解析センター	ヒトゲノム解析センター長 宮野 悟