

村田 健史 (情報通信研究機構)

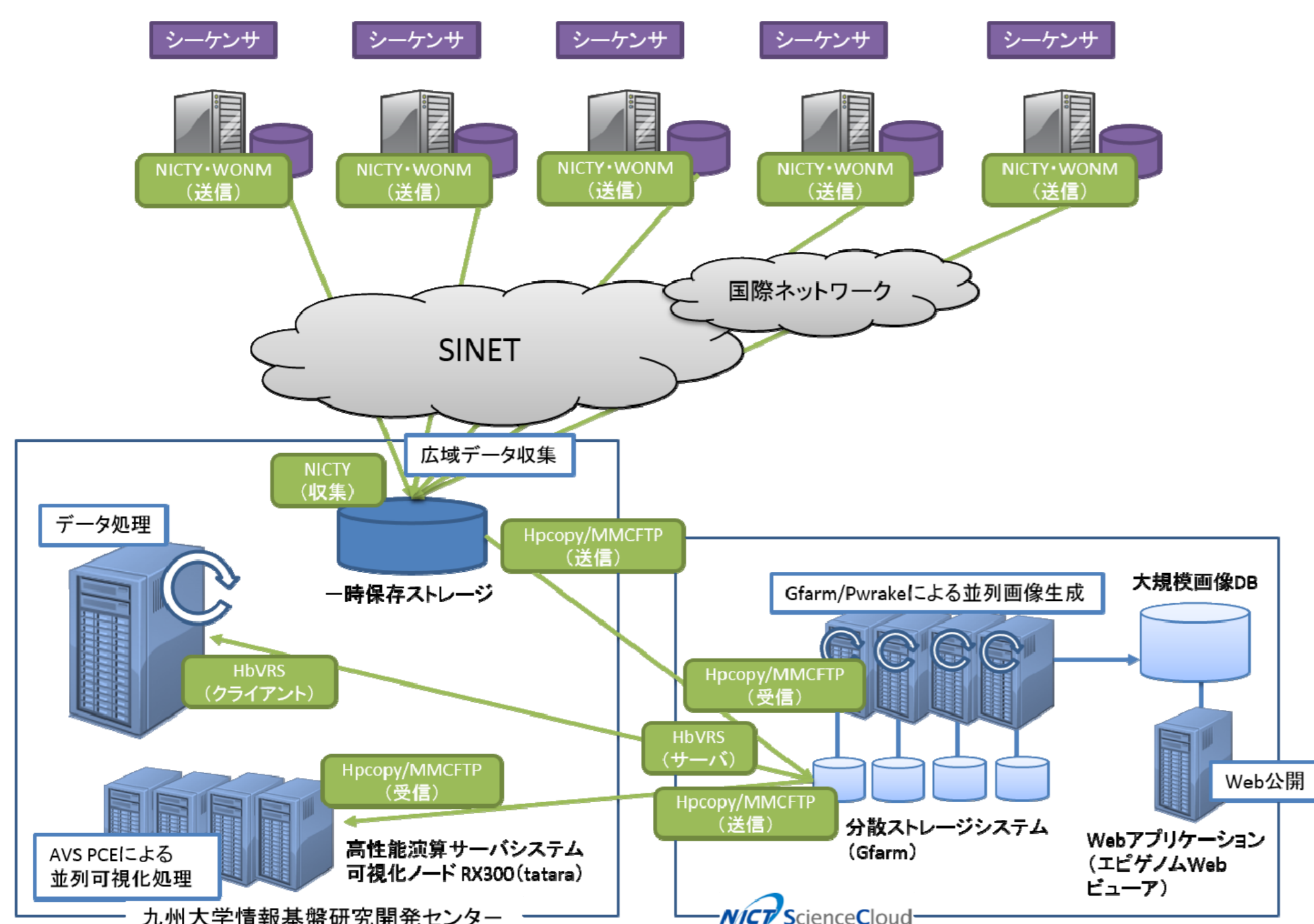
# SINETを活用したエピゲノムビッグデータ可視化システム技術開発



本申請では、今後ますますビッグデータ化が進むエピゲノム研究において、九州大学と情報通信研究機構をネットワーク(SINET)で接続し、仮想的な大規模ラボ(以下、「バーチャルエピゲノムラボ」と呼ぶ。)によるエピゲノム研究を行うための基盤技術開発を行う。バーチャルエピゲノムラボを実現するため、情報通信研究機構及び情報学研究所によるビッグデータ要素基盤技術を選材適所で活用する。具体的期には、左図に示すようなシステムの実現を目指す。

- ① シーケンサデータ収集: ビッグデータエピゲノム研究では、国内外の様々なシーケンサデータを収集し、ヒトまたはその他の生物の様々な部位におけるゲノム情報を融合的に解析することが求められる。ここでは、情報通信研究機構が開発したグローバルなデータファイルクロージングシステム(NICTYおよびWONM)を使うことで、運用者に負担にならずにデータをクロージングする。ただし、データ公開については研究機関ごとのポリシーが異なるため、左図のようにNICTYの実装を許可されない組織も想定される。これらについては、対象機関ごとに個別交渉を行うこととする。なお、データサイズが大きい場合には、データ通信プロトコルにHTTPやFTPではなく、後述のHpFPやMMCFTPを用いることもある。
- ② データ伝送: 九州大学にキャッシュされたデータファイルは、適宜、情報通信研究機構のNICTサイエンスクラウドに転送される。転送データサイズが大きい場合、HpFPやMMCFTPを用いて高速データ伝送を行う。HpFPは情報通信研究機構が開発した独自のデータ通信プロトコルであり、ファイル転送ツールとしてhpcopyがある。MMCFTPは情報学研究所が開発した並列データ伝送ツール(アプリケーション)である。データタイプやネットワークに応じて、より有効なツールを用いることとする。伝送されたデータファイルは、情報通信研究機構の分散ファイルシステム(Gfarm)に保存される。これらのプロトコルまたはデータ伝送ツールにより、九州大学と情報通信研究機構間で最大10Gbps(通常はファイアウォールがあるために1Gbps)でのファイル伝送を実現する。
- ③ データ可視化: 情報通信研究機構が開発したSTARStouchにより、30億にもなる塩基列上にエピゲノム情報をスケラブルに可視化する。想定される可視化データファイルはデータごとに数100万から数1000万と予想されるため、並列可視化が必要である。並列可視化は、九州大学情報基盤研究開発センターの可視化サーバ(RX300)およびNICTサイエンスクラウドの並列サーバを用いる。前者においてはAVS PCEを用いた並列可視化を、後者においてはPwrakeを用いた並列可視化を実現する。STARStouch技術によるエピゲノムビューアはプロトタイプ開発が終わり、今後はアニメーションの追加やUIの変更、データの追加を行う。
- ④ 可視化公開: STARStouchによる大規模可視化WebをNICTサイエンスクラウド上において行う。ここでは、特にWebブラウジングの高速化を進める。Webブラウザの高速化では、Google社のQUICに対抗して情報通信研究機構が開発したHpFPプロトコルハンドラをWebサーバに組み込む。これにより、IE(Internet Explorer)限定ではあるが、通常のHTTPに比較して数倍の高速が期待される。動作検証は、九州大学側のPCからアクセスすることで行う。

なお、本研究開発は、JHPCN2015(「クラウドを活用したビッグデータポスト処理環境実現のためのデータ伝送実験」、代表者:村田)の成果をゲノム分野での実利用に発展・実用化させるテーマである。JHPCN2015ではSINETによる組織間データ伝送の基盤技術を確認したが、本申請ではこの開発技術を九州大学の先端的研究分野の一つであるエピゲノム研究に実用化(データ収集、データ伝送、データ処理、データ可視化)することを目指している。なお、本申請内容はJHPCN2016には申請していない。これは、申請締め切り段階では代表者のH.28年度(情報通信研究機構ではH.28年度が新しい中長期計画の1年目となる)の所属および研究継続の可能性が不明であった(端的に言うと、申請者が研究職を継続できる保証がなかった)ために申請を断念したことによる。しかし、一方でJHPCN2015の成果は着実に挙がっており、評価者からも「これらの技術を具体的な科学研究テーマに適用することが望ましい」という指摘を受けている。申請段階(2016年3月末)において代表者の次年度(次期中長期計画)での研究職が確定したため、本申請を行ったものである。



ビッグデータエピゲノムシステム概要



STARStouchによるエピゲノム可視化(～H.27の成果): 横方向は塩基配列(1000～30億までをスケラブルに表示・縦方向は細胞組織を表す。縦方向は任意の順序に並べ替えることができ、特定の複数のデータをオーバーレイ表示することができる。

## H.28年度業績(査読付き国際学会論文)

1. Murata, K. T., Pavarangkoon, P., Yamamoto, K., Nagaya, Y., Katayama, N., Muranaga, K., Mizuhara, T., Takaki, A. and Kimura, E., An Application of Novel Communications Protocol to High Throughput Satellites, The 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON2016), Vancouver, Canada, Oct. 13-15, 2016.
2. Murata, K. T., Muranaga, K., Yamamoto, K., Nagaya, Y., Pavarangkoon, P., Satoh, S., Mizuhara, T., Kimura, E., Tatebe, O., Tanaka, M. and Kawahara, S., Real-time 3D Visualization of Phased Array Weather Radar Data via Concurrent Processing in Science Cloud, The 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON2016), Vancouver, Canada, Oct. 13-15, 2016.
3. Murata, K. T., Pavarangkoon, P., Yamamoto, K., Nagaya, Y., Muranaga, K., Mizuhara, T., Takaki, A., Tatebe, O. and Kimura, E., Multiple Streams of UDT and HpFP Protocols for High-bandwidth Remote Storage System in Long Fat Network, The 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON2016), Vancouver, Canada, Oct. 13-15, 2016.
4. Pavarangkoon, P., Murata, K. T., Okada, M., Yamamoto, K., Nagaya, Y., Mizuhara, T., Takaki, A., Muranaga, K. and Kimura, E., Bandwidth Utilization Enhancement for INTEL SAT Satellite Network, The 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON2016), Vancouver, Canada, Oct. 13-15, 2016.
5. Murata, K. T., Pavarangkoon, P., Suzuki, K., Yamamoto, K., Nagaya, A., Kan T., Katayama N., Yahata M., Muranaga, K., Mizuhara, T., Takaki, A. and Kimura, E., A High-Speed Data Transfer Protocol for Geostationary Orbit Satellites, 2016 International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, Oct. 12-14, 2016.
6. Murata, K. T., Pavarangkoon, P., Yamamoto, K., Nagaya, Y., Satoh, S., Muranaga, K., Mizuhara, T., Takaki, A. and Kimura, E., Improvement of Real-time Transfer of Phased Array Weather Radar Data on Long-Distance Networks, The 2016 International Conference on Radar, Antenna, Microwave, Electronics and Telecommunications (2016ICRAMET), Jakarta, Indonesia, Oct. 3-5, 2016.
7. K. T. Murata, P. Pavarangkoon, K. Yamamoto, Y. Nagaya, T. Mizuhara, A. Takaki, K. Muranaga, E. Kimura, T. Ikeda, K. Ikeda, and J. Tanaka, A Quality Measurement Tool for High-Speed Data Transfer in Long Fat Networks, 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM2016), Split, Croatia, Sep. 22-24, 2016.

大川 恭行・教授・九州大学 生体防御医学研究所 トランスオミクス医学研究センター・yohkawa@bioreg.kyushu-u.ac.jp  
 前原 一満・助教・九州大学 生体防御医学研究所 トランスオミクス医学研究センター・kazumits@bioreg.kyushu-u.ac.jp  
 南里 豪志・准教授・九州大学 情報基盤研究開発センター 先端計算基盤研究部門・nanri@cc.kyushu-u.ac.jp  
 長屋 嘉明・総務省 国際政策課  
 山本 和憲・情報通信研究機構 総合テストベッド研究開発推進センター・kaz-y@nict.go.jp  
 深沢 圭一郎・准教授・京都大学 学術情報メディアセンター コンピューティング研究部門・fukazawa@media.kyoto-u.ac.jp  
 山中 顕次郎・特任准教授・国立情報学研究所 先端ICTセンター・yamanaka@nii.ac.jp  
 水原 隆道・代表取締役・株式会社クリアリンクテクノロジー・mizuhara@clealink.jp  
 村田 健史・研究統括・情報通信研究機構 総合テストベッド研究開発推進センター・ken.murata@nict.go.jp