

課題名	1 分子長鎖再解読に基づく標準マウスゲノム配列および構造決定と公開																						
課題管理者	榎藤 洋一 国立研究開発法人理化学研究所																						
実施期間	2015 年度																						
概要・実施体制	<p>ヒトゲノム解読に続き、マウスゲノムが他の高等動物に先立って解読されました。これまで、多くのマウス近交系ゲノムが標準参照配列として決定されたものの、公開されている公共ゲノムデータベースからも明らかのようにまだまだ多くの未解読配列（ギャップ）が存在します（図参照）。ヒトゲノムでは 2015 年に参照配列の見直しが進み、ギャップの解消とともに逆位転座重複などの高次ゲノム構造の是正が急速に進んでいます。ヒト疾患はじめモデル系として広く利用されているマウスゲノムにおいても同様の抜本的な見直しが喫緊の課題となっています。そこで、理研バイオリソースセンターが企画し、ゲノム支援を受け大規模ゲノム解析および突然変異解析を進めているマウスゲノムビッグデータを基盤として、最新の技術による 1 分子レベルで 10kb を越える長鎖 DNA 解読データを加味することで、未解読領域の解消、高次ゲノム構造の見直し、および、高精度塩基解読を整備し公開します。とくにオリジナルデータから公開することにより、さらにさまざまなデータと組み合わせ、公開データ改善やアノテーションの追加が誰でもいつでも実施することが可能となるプラットフォーム構築の基盤整備を目指します。</p> <table border="1" data-bbox="352 725 1145 1249"> <thead> <tr> <th data-bbox="352 725 687 770">マウスゲノム参照配列情報</th> <th data-bbox="938 725 1114 770">実験用マウス</th> </tr> </thead> <tbody> <tr> <td data-bbox="352 779 820 824">NCBI/Genome Reference Consortium</td> <td data-bbox="948 779 1104 824">GRCm38.p4</td> </tr> <tr> <td data-bbox="352 853 655 887">Total Bases in Assembly</td> <td data-bbox="948 853 1126 887">2,803,568,840</td> </tr> <tr> <td data-bbox="352 893 743 927">Total Non-N Bases in Assembly</td> <td data-bbox="948 893 1126 927">2,724,212,084</td> </tr> <tr> <td data-bbox="352 934 676 967">Total assembly gap length</td> <td data-bbox="986 934 1126 967">79,356,756</td> </tr> <tr> <td data-bbox="352 974 651 1008">Gaps between scaffolds</td> <td data-bbox="1075 974 1126 1008">191</td> </tr> <tr> <td data-bbox="352 1014 603 1048">Number of scaffolds</td> <td data-bbox="1075 1014 1126 1048">293</td> </tr> <tr> <td data-bbox="352 1055 517 1088">Scaffold N50</td> <td data-bbox="986 1055 1126 1088">52,589,046</td> </tr> <tr> <td data-bbox="352 1095 580 1128">Number of contigs</td> <td data-bbox="1075 1095 1126 1128">796</td> </tr> <tr> <td data-bbox="352 1135 496 1169">Contig N50</td> <td data-bbox="986 1135 1126 1169">32,273,079</td> </tr> <tr> <td data-bbox="352 1176 895 1209">Total number of chromosomes and plasmids</td> <td data-bbox="1091 1176 1126 1209">22</td> </tr> </tbody> </table> <p data-bbox="352 1218 1134 1249">2015年10月26日現在、完全ギャップが191箇所、Nで補完された未解読配列が80Mb残っている。</p>	マウスゲノム参照配列情報	実験用マウス	NCBI/Genome Reference Consortium	GRCm38.p4	Total Bases in Assembly	2,803,568,840	Total Non-N Bases in Assembly	2,724,212,084	Total assembly gap length	79,356,756	Gaps between scaffolds	191	Number of scaffolds	293	Scaffold N50	52,589,046	Number of contigs	796	Contig N50	32,273,079	Total number of chromosomes and plasmids	22
マウスゲノム参照配列情報	実験用マウス																						
NCBI/Genome Reference Consortium	GRCm38.p4																						
Total Bases in Assembly	2,803,568,840																						
Total Non-N Bases in Assembly	2,724,212,084																						
Total assembly gap length	79,356,756																						
Gaps between scaffolds	191																						
Number of scaffolds	293																						
Scaffold N50	52,589,046																						
Number of contigs	796																						
Contig N50	32,273,079																						
Total number of chromosomes and plasmids	22																						
成果	<p>(1) ショートリードビッグデータに基づく高精度塩基配列決定公開されている C57BL/6J マウスゲノム由来の標準参照配列 GRCm38p4 (mm10) と比較検討した結果、本邦で広く利用されている C57BL/6Jdcl 標準近交系マウスのゲノム配列では、少なくとも 4601 箇所の塩基配列が異なっていることを確認した。また、この解析の過程で、広く遺伝的解析などに用いられる公開マウス dbSNP 情報のうち少なくとも 1 万個の SNP 情報は間違っていることを強く示す解析結果を得た。</p> <p>(2) 1 分子長鎖リードのための諸条件の検討さまざまな臓器から抽出したゲノム DNA をパルスフィールド電気泳動法を用いて比較検討し、超高分子ゲノム DNA の調製には、肺ゲノムが最も適することを確認した。また、平均 10kb 長で 10×カバレッジの解読するためには、PacBioRSII の 6 時間モードが最も適していると総合的に判断し、そのためのライブラリー作製およびシーケンシングを行った。</p> <p>(3) マウスゲノムの 1 分子長鎖解読 最終的に 7,519,270 の総リード数を得た。解読塩基数はのべ 96,522,590,619bp と約 30×カバレッジのビッグデータを得ることができ当初目的を大きく越えた。また平均リード長も 12,837bp と当初目的の平均 10kb を上回った値が得られた。</p> <p>(4) 成果公開のための整備 DDBJ からオリジナルデータを公開するための登録 ID を取得した。また総合的にアセンブリーしたデータを公開するため、理研情報基盤センターが整備している理研メタデータベース上に公開用プラットフォームを構築した。本研究による成果と公開により、疾患モデル開発における責任遺伝子や原因変異同定の精度が飛躍的にあがる。さらに、ゲノム参照配列が基盤となる一般的な PCR 解析やゲノム編集の精度向上にも大きく寄与し、モデル動物としてのマウスリソースの価値が高まる。</p> <p>【成果の発信】</p>																						

- | | |
|--|---|
| | <ol style="list-style-type: none">1. "Detection of spontaneous mutations in mammalian genomes." Y. Gondo in "Biological & Medical Science based on Physics: Radiation and Physics, Physics on Medical Science, Modeling for Biological System" Kyoto, 2015/11/5, 国際会議2. "Needs of fundamental revision of mouse genome reference sequences" Y. Gondo et al. in "29th International Mammalian Genome Conference (IMGC2015)" Yokohama, 2015/11/10, 国際会議3. 「標準マウス系統 C57BL/6J における自然発生突然変異の解析法の確立」権藤洋一ほか、第 38 回日本分子生物学会/第 88 回日本生化学会合同大会、神戸市、2015/12/1, 国内会議4. "Genetics, mutagenesis and genomics with standardized mouse inbred strains" Y. Gondo at School of Pharmacy and School of Life Sciences, Fudan University, Shanghai, China, 2015/12/17, 海外招待講演5. "Accumulation and detection of germline spontaneous mutations in C57BL/6Jcl inbred mouse strain" Y. Gondo et al. in "30th International Mammalian Genome Conference (IMGC2015)" Orland, USA, 2016/7/14, 国際会議 |
|--|---|