

システム設計(9/10)

【類似性検索機能の画面設計案(1/2)】

- 使用する遺伝子セットの数は計算時間に線形に効いてくる。デフォルトで幾つかはセットしておくものの、ユーザに選択させるのがよいと考えられる。

問い合わせの登録

検索対象DBの選択

●全部 ○ヒト ○マウス ○ラット

Gene setの選択

MSigDB

| | |
|------------------------------|--|
| EXTRINSIC_TO_PLASMA_MEMBRANE | |
| ORGANELLE_PART | |
| CELL_PROJECTION_PART | |
| CYTOPLASMIC_VESICLE_MEMBRANE | |

計算に使用されるもの(100まで)

| | |
|--------------------------------|--|
| ALL | |
| HSIAO_HOUSEKEEPING_GENES | |
| MIKKELSEN_PLURIPOTENT_STATE_UP | |
| MIKKELSEN_PLURIPOTENT_STATE_DN | |

ドラッグでカードを移動させるイメージ

Copyright (C) Mitsubishi Research Institute, Inc.

17

システム設計(10/10)

【類似性検索機能の画面設計案(2/2)】

- 基本は、前項で示したテーブル。絞り込み、ソート機能を充実させる。

GSM503052を問い合わせとした検索結果

| 問い合わせ | DB | 遺伝子数 | 相関係数 | p値 | 遺伝子セット |
|-----------|-----------|------|----------|----------|---|
| GSM603052 | GSM603054 | 146 | 0.995229 | 0.00E+00 | PEGE, MAMMARY, STEM CELL, DN |
| GSM603052 | GSM603054 | 388 | 0.994772 | 0.00E+00 | HSIAO_HOUSEKEEPING_GENES |
| GSM603052 | GSM603050 | 146 | 0.994443 | 0.00E+00 | PEGE, MAMMARY, STEM CELL, DN |
| GSM603052 | GSM603054 | 33 | 0.994076 | 0.00E+00 | VANOVA, HEMATOPOIESIS, STEM CELL, SHORT TERM |
| GSM603052 | GSM603054 | 206 | 0.993372 | 0.00E+00 | RAMALHO, STEMNESS UP |
| GSM603052 | GSM603050 | 206 | 0.991623 | 0.00E+00 | RAMALHO, STEMNESS UP |
| GSM603052 | GSM603054 | 888 | 0.991619 | 0.00E+00 | REACTOME, IMMUNE SYSTEM |
| GSM603052 | GSM603050 | 888 | 0.991073 | 0.00E+00 | REACTOME, IMMUNE SYSTEM |
| GSM603052 | GSM603054 | 274 | 0.990976 | 0.00E+00 | OSWALD, HEMATOPOIETIC, STEM CELL, IN COLLAGEN GEL, UP |
| GSM603052 | GSM603050 | 123 | 0.990989 | 0.00E+00 | KEGG, SYSTEMIC LUPUS ERYTHEMATOSUS |
| GSM603052 | GSM603054 | 76 | 0.990996 | 0.00E+00 | KEGG, PHOSPHATIDYLINOSITOL, SIGNALING |
| GSM603052 | GSM603054 | 24 | 0.990937 | 0.00E+00 | BYSTRYKH, HEMATOPOIESIS, STEM CELL, IN COLLAGEN GEL, DN |
| GSM603052 | GSM603050 | 220 | 0.988964 | 0.00E+00 | ENDOMEMBRANE, SYSTEM |
| GSM603052 | GSM603054 | 38 | 0.988933 | 0.00E+00 | LIANG, HEMATOPOIESIS, STEM CELL, IN COLLAGEN GEL, DN |
| GSM603052 | GSM603050 | 874 | 0.988308 | 0.00E+00 | BYSTRYKH, HEMATOPOIESIS, STEM CELL, IN COLLAGEN GEL, DN |
| GSM603052 | GSM603050 | 66 | 0.988033 | 0.00E+00 | BEIER, GLIOMA, STEM CELL, DN |
| GSM603052 | GSM603054 | 316 | 0.987877 | 0.00E+00 | JAAITINEN, HEMATOPOIETIC, STEM CELL, UP |
| GSM603052 | GSM603054 | 2199 | 0.987286 | 0.00E+00 | ALL |
| GSM603052 | GSM603050 | 2199 | 0.987103 | 0.00E+00 | ALL |
| GSM603052 | GSM603050 | 672 | 0.987043 | 0.00E+00 | VANOVA, HEMATOPOIESIS, STEM CELL, AND PROGENITOR |
| GSM603052 | GSM603054 | 94 | 0.986888 | 0.00E+00 | HOEBEKE, LYMPHOID, STEM CELL, UP |
| GSM603052 | GSM603050 | 332 | 0.986709 | 0.00E+00 | IMMUNE SYSTEM, PROCESS |
| GSM603052 | GSM603054 | 232 | 0.986482 | 0.00E+00 | OSWALD, HEMATOPOIETIC, STEM CELL, IN COLLAGEN GEL, UP |
| GSM603052 | GSM603054 | 716 | 0.98594 | 0.00E+00 | WONG, ADULT TISSUE, STEM MODULE |
| GSM603052 | GSM603054 | 299 | 0.985134 | 0.00E+00 | VANOVA, HEMATOPOIESIS, STEM CELL, LONG TERM |
| GSM603052 | GSM603050 | 274 | 0.984803 | 0.00E+00 | OSWALD, HEMATOPOIETIC, STEM CELL, IN COLLAGEN GEL, DN |
| GSM603052 | GSM603050 | 88 | 0.983894 | 0.00E+00 | HOEBEKE, LYMPHOID, STEM CELL, DN |

表示されるヒットを絞り込む機能 (相関係数、p値に対する閾値に対する絞り込み機能があってもよい)

ヘッダー部分クリックでソート (相関係数もしくはp値)

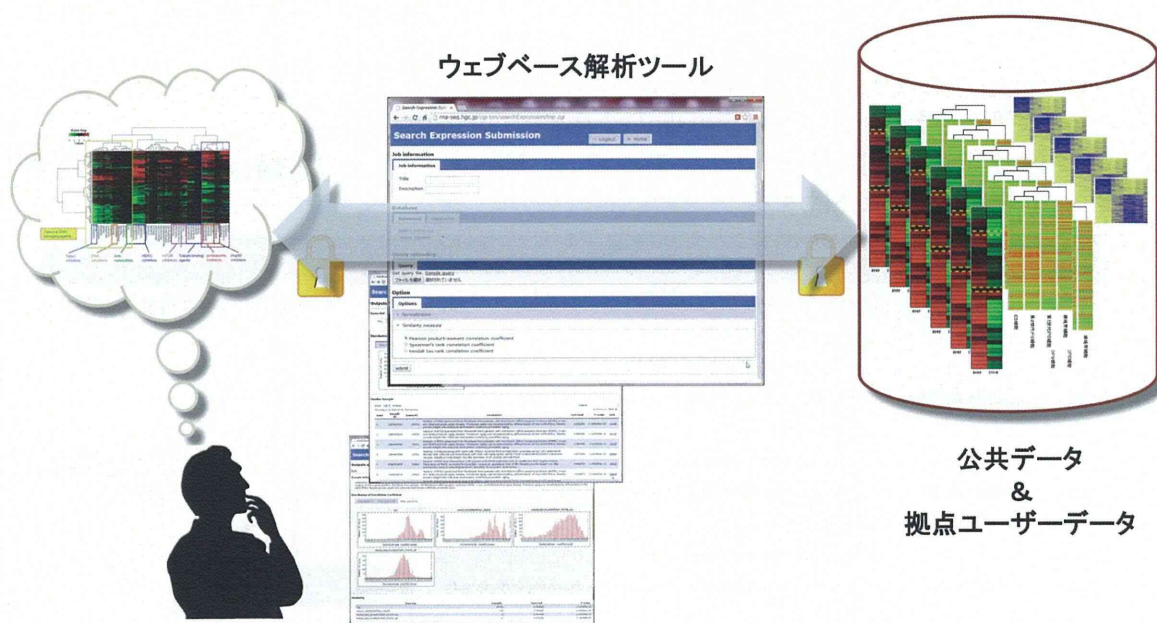
Copyright (C) Mitsubishi Research Institute, Inc.

18

厚生労働科学研究費補助金
難病・がん等の疾患分野の医療の実用化研究事業
iPS細胞、ES細胞、体性幹細胞の解析ツールへの機能追加

株式会社三菱総合研究所

Copyright (C) Mitsubishi Research Institute, Inc.



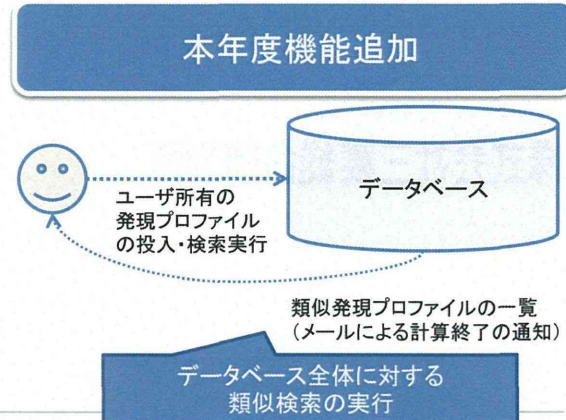
Copyright (C) Mitsubishi Research Institute, Inc.

2

遺伝子発現プロファイル検索システムに対する機能追加(1/2)

【操作性の簡便化】

- 手持ちの発現プロファイルデータを問い合わせとしてシステムに与え、簡単なパラメータ設定を行いサブミットするだけで、データベース全体に対する検索が行えるように改修する
 - ・ キーワード検索による検索対象の絞り込みを廃止
 - ・ 長時間計算ジョブ・同時検索要求を適切に処理できるようにシステムを改修
 - ・ 計算結果が確実にユーザに届けられる仕組みを実装



Copyright (C) Mitsubishi Research Institute, Inc.

3

遺伝子発現プロファイル検索システムに対する機能追加(2/2)

- 検索結果に対する付加情報の追加
 - ・ 検索結果として得られる類似発現プロファイル一覧に細胞種や実験条件などのアノテーションを追加する
 - ・ 発現プロファイルがどのように似ているのかを、既知遺伝子セットでの類似度で確認できるようにする(MSigDBの活用)
 - ・ 類似検索に問題がないか(次元の呪いに陥っていないか)を類似スコアの分布で可視化できるようにする

| Rank | Sample ID | Annotation | Corr. Coef. | P-value | Link |
|------|-----------|--|-------------|---------|------|
| 1 | GSM189751 | Intestinal mucosa biopsy from healthy human individuals other: pair 01 gender: female age: 30 genotype/variation: monozygotic twin | 1.000 | | |
| 2 | GSM189754 | Intestinal mucosa biopsy from healthy human individuals other: pair 02 gender: female age: 39 genotype/variation: monozygotic twin | 0.998 | | |
| 3 | GSM189753 | Intestinal mucosa biopsy from healthy human individuals other: pair 02 gender: female age: 39 genotype/variation: monozygotic twin | 0.998 | | |
| 4 | GSM189763 | Intestinal mucosa biopsy from healthy human individuals other: pair 07 gender: female age: 39 genotype/variation: monozygotic twin | | | |

類似スコアの度数分布のグラフ

類似スコアの度数分布のグラフ

類似度

遺伝子数:2050

MSigDB
Molecular Signatures Database

類似度

遺伝子数:15040

Copyright (C) Mitsubishi Research Institute, Inc.

4

問い合わせデータ投入ページ

- 問い合わせとなる発現プロファイルデータのアップロード設定
- 検索対象となるデータベースの選択・検索パラメータの設定
- サブミットボタンを押下

⇒データがアップロード・データベース検索開始

データベース選択
(GEOから抽出されたヒト、マウス、ラットの発現プロファイルデータ)

MSigDBの遺伝子セットの選択
(ここで選択された遺伝子セットで類似度が計算される)

問い合わせデータアップロード

検索パラメータの設定
(類似度を測る相関係数(Pearson等)、正規化法)

サブミットボタン

Copyright (C) Mitsubishi Research Institute, Inc.

5

計算結果ページ(一覧)

遺伝子セット選択ボタン
(ALLボタンは全遺伝子を使用しての相関係数計算結果を表示)

相関係数の度数分布

ヒット一覧
(ヒットしたサンプルのID、アノテーション、相関係数計算に使用された遺伝子数、相関係数、p値を一覧表示)

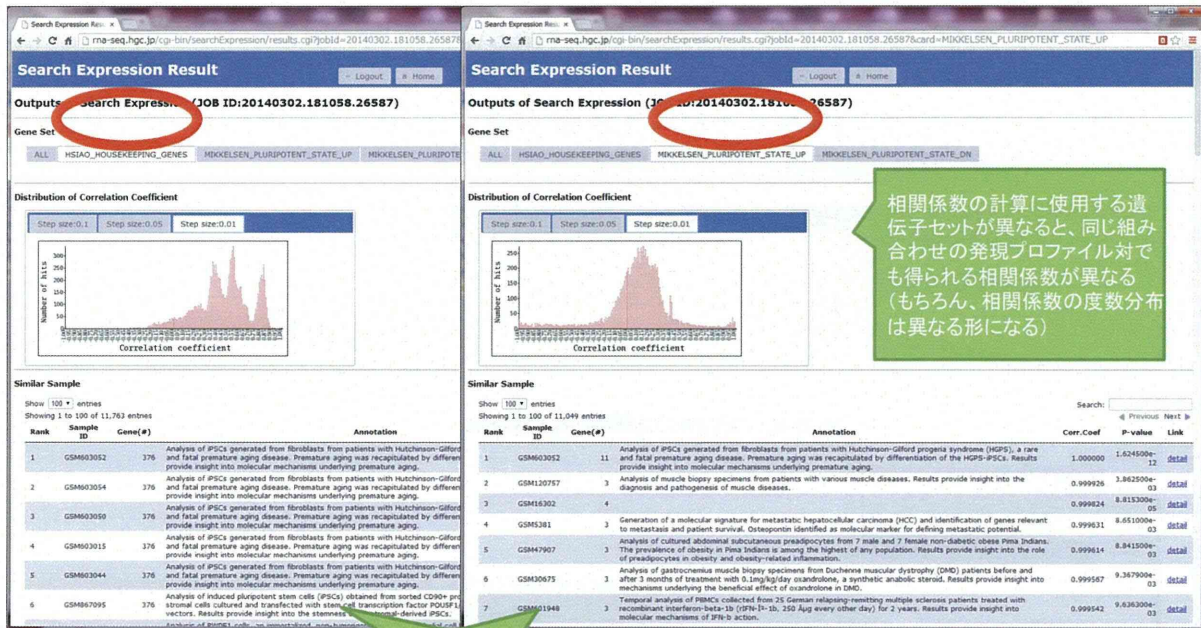
計算結果(詳細)へのリンク

| Rank | Sample ID | Gene(#) | Annotation | Corr.Coeff | P-value | Link |
|------|-----------|---------|--|------------|--------------|------------------------|
| 1 | GSM603052 | 20401 | Analysis of iPSCs generated from fibroblasts from patients with Hutchinson-Gilford progeria syndrome (HGPS), a rare and fatal premature aging disease. Premature aging was recapitulated by differentiation of the HGPS-iPSCs. Results provide insight into molecular mechanisms underlying premature aging. | 1.000000 | 0.000000e+00 | detail |
| 2 | GSM603054 | 20401 | Analysis of iPSCs generated from fibroblasts from patients with Hutchinson-Gilford progeria syndrome (HGPS), a rare and fatal premature aging disease. Premature aging was recapitulated by differentiation of the HGPS-iPSCs. Results provide insight into molecular mechanisms underlying premature aging. | 0.983580 | 1.624500e-12 | detail |
| 3 | GSM603050 | 20401 | Analysis of iPSCs generated from fibroblasts from patients with Hutchinson-Gilford progeria syndrome (HGPS), a rare and fatal premature aging disease. Premature aging was recapitulated by differentiation of the HGPS-iPSCs. Results provide insight into molecular mechanisms underlying premature aging. | 0.983365 | 1.624500e-12 | detail |
| 4 | GSM867095 | 20401 | Analysis of induced pluripotent stem cells (iPSCs) obtained from sorted CD90+ prostate cancer (PC)-associated stromal cells cultured and transfected with stem cell transcription factor POUSF1/LIN28/NANOG/SOX2 expression vectors. Results provide insight into the stemness of PC stromal-derived iPSCs. | 0.872203 | 1.624400e-12 | detail |
| 5 | GSM763979 | 20401 | Analysis of iPSC lines derived from a PD patient with SNCA triplication and an unaffected first-degree relative. Triplication of SNCA, encoding α -synuclein, causes an aggressive form of PD. Results provide insight into the mechanistic basis of neurodegeneration caused by α -synuclein dysfunction. | 0.840528 | 1.624500e-12 | detail |
| 6 | GSM603015 | 20401 | Analysis of iPSCs generated from fibroblasts from patients with Hutchinson-Gilford progeria syndrome (HGPS), a rare and fatal premature aging disease. Premature aging was recapitulated by differentiation of the HGPS-iPSCs. Results provide insight into molecular mechanisms underlying premature aging. | 0.839473 | 1.624600e-12 | detail |

Copyright (C) Mitsubishi Research Institute, Inc.

6

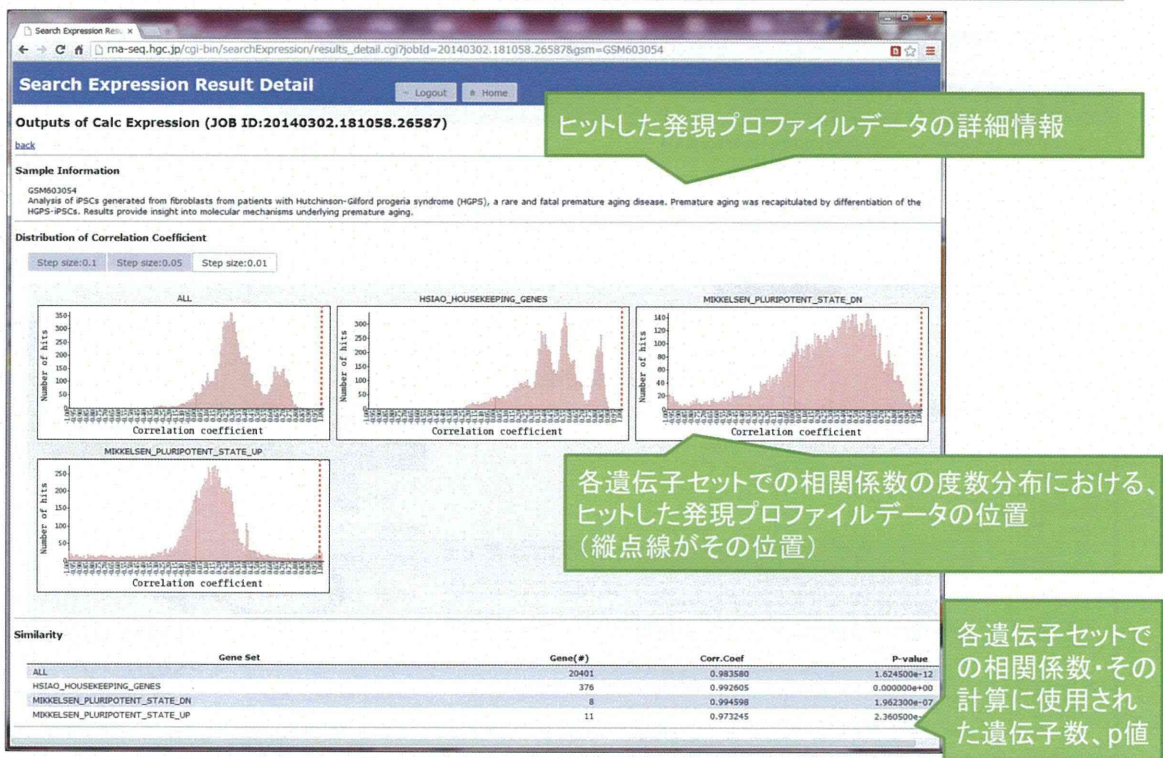
計算結果ページ(一覧): 遺伝子セットの比較



Copyright (C) Mitsubishi Research Institute, Inc.

7

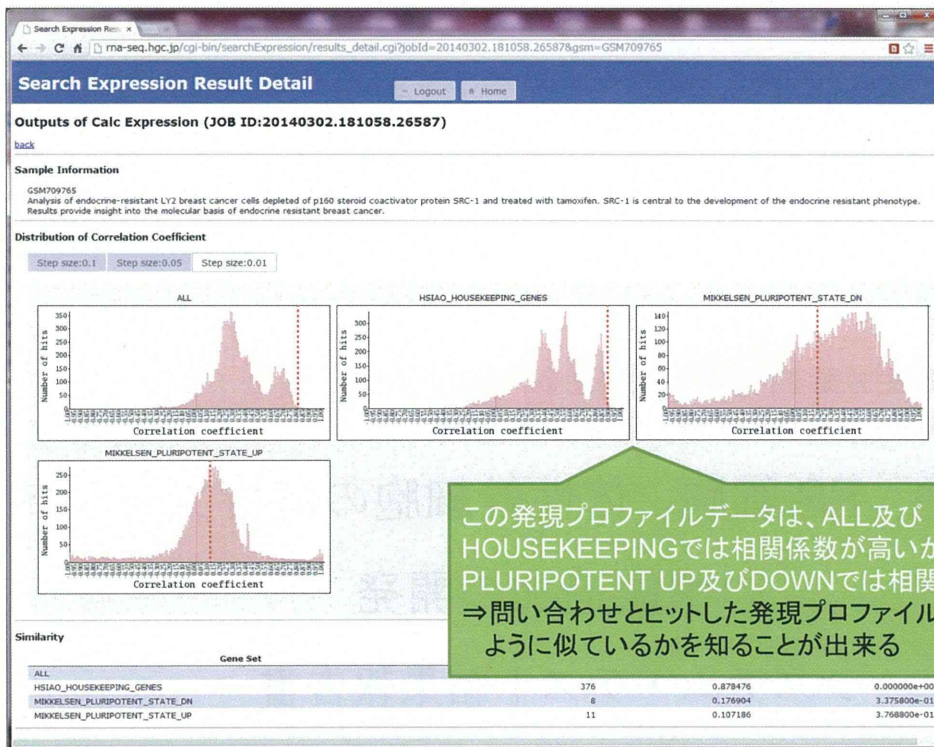
計算結果ページ(詳細)



Copyright (C) Mitsubishi Research Institute, Inc.

8

計算結果ページ(詳細): 遺伝子セットの比較



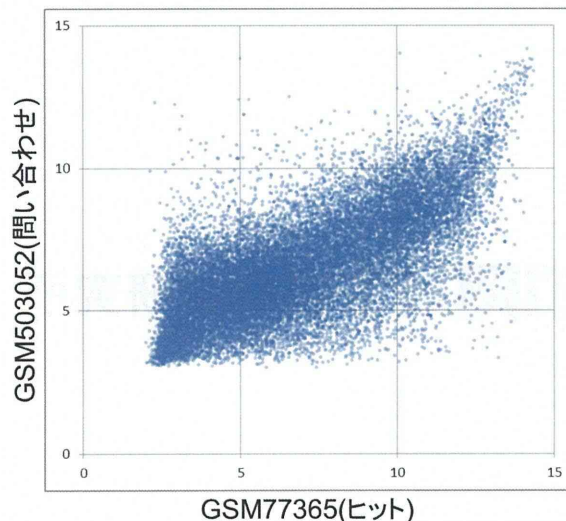
Copyright (C) Mitsubishi Research Institute, Inc.

9

課題

発現プロファイルデータ対がどのように似ているか、逆にどの遺伝子の発現が異なっているかの情報を提供できないか

- Scatter plot
- 高発現・低発現遺伝子の一覧表示
- 発現レベルが異なる遺伝子の抽出(Differential gene expression)



Copyright (C) Mitsubishi Research Institute, Inc.


10

平成 25 年度厚生労働省科学研究費補助金
難病・がん等の疾患分野の医療の実用化研究事業

iPS 細胞、ES 細胞、体性幹細胞の解析ツールへ の機能追加の開発

—システム・テスト結果報告書—

2014 年 3 月 25 日

 株式
会社 三菱総合研究所

以下に、テスト結果を示す。

| テスト番号 | テスト内容 | 期待される結果 | 結果 |
|-------|---|--|----|
| 1 | テストデータのアップロード・ジョブ投入 | テストデータを格納したファイルが存在する際、その内容がアップロードされ、計算終了待ち受けページに遷移する 計算終了後、計算結果ページに遷移する | OK |
| 1.1 | 正テストデータがアップロード・ジョブ投入された場合 | 計算結果ページにヒット一覧が表示される | OK |
| 1.2 | 誤テストデータがアップロード・ジョブ投入された場合 | 計算結果ページにヒットが表示されない | OK |
| 1.3 | ジョブタイトル・説明文章が入力され・ジョブ投入された場合 | 計算結果ページにジョブタイトル・説明文章が表示される | OK |
| 1.4 | 1.1の場合に、遺伝子セットを選択され・ジョブ投入された場合 | 選択された遺伝子セットの相関係数が計算され、計算結果ページにそのヒット一覧と度数分布プロットが表示される | OK |
| 1.5 | 1.1の場合に、正規化法としてTMMが選択され・ジョブ投入された場合 | TMMによる正規化を受けた発現値が、サーバ内のジョブディレクトリに保存される（管理者の立場として確認） | OK |
| 1.6 | 1.1の場合に、デフォルトとは異なる相関係数の種類（スピアマンの順位相関係数及びケンドールの順位相関係数）が選択され・ジョブ投入された場合 | デフォルトとは異なるヒット、相関係数の度数分布プロットが得られる | OK |
| 2 | 1.1の場合に、ジョブ終了のメール通知を受ける | メールに示されたURLに従ってヒット一覧ページにアクセスする | OK |
| 3 | ジョブ一覧ページの正常終了しているジョブを選択する | 該当ジョブのヒット一覧ページにアクセスする | OK |

| テスト 番号 | テスト内容 | 期待される結果 | 結果 |
|-----------|--|----------------------|----|
| 4 | 1.1の場合に、ヒット一覧ページからヒット詳細ページに遷移する | ヒット詳細ページで期待される表示を得る | OK |
| 5 | 1.1の場合に、ヒット詳細ページから遺伝子一覧ページに遷移する | 遺伝子一覧ページで期待される表示を得る | OK |
| 6 | 1.1の場合に、遺伝子一覧ページにおいて、散布図の軸を対数軸に変換するアクションを加える | 散布図の指定された軸が対数軸に変更される | OK |

平成 25 年度厚生労働省科学研究費補助金
難病・がん等の疾患分野の医療の実用化研究事業
タブレットを用いた再生医療に関する
標準作業手順書登録・閲覧・入力システムの開発

仕様書

平成 25 年 11 月

東京大学医科学研究所