Vol.2010-BIO-20 No.9
2010/3/5

正則化正準相関解析を用いた抗がん剤の影響による共通パスウェイ解析

金蕙鈴[1,2], 加藤毅[3,2], 茂櫛薫[1],
田中博[1], 藤渕航[2,1]

患者の QOL(Quality Of Life)における重要性にも関らず，薬物に由来する作用メカニズムは不明な点が多く，これまでにシステム的なアプローチが極端に少なかった． 我々は，汎化能力の高いリッジパラメータを導入した正則化正準相関解析を行い，遺伝子活性とパスウェイの相関関係を解析した．その結果，細胞周期関連パスウェイにおいて薬理学的または化学的の分類の異なる抗がん剤同士の関連性を見出すことができた．

Detection of common pathways activated by anticancer drugs using regularized canonical correlation analysis

Hyeryung Kim[1,2], Tsuyoshi Kato[3,2], Kaoru Mogushi[1],
Hiroshi Tanaka[1] and Wataru Fujibuchi[2,1]

In spite of the importance to improve the quality of life (QOL) of patients, few systematic approaches have been available to uncover molecular mechanisms of drug-action. We adopted a new approach to solve this problem using the regularized canonical correlation analysis which is expected to enhance generalizing capability of finding correlations between gene activities and pathway information. As a result, we found a novel relationship between various anti-cancer drugs that are classified in different pharmacological or chemical groups.

1. はじめに

近年のゲノム分野における様々な技術の発展によって，膨大な生物データが蓄積されてきた．DNA チップを用いた遺伝子発現データをはじめ，代謝およびシグナル伝達に関するパスウェイデータ，ChIP-chip データ等がその例である．これらのデータは生体分子に関する情報であることは共通であるものの，転写産物の量，代謝する化合物または遺伝子間相互作用の有無，または DNA と転写因子との相互作用を表すなど，データの内容やプラットフォームは様々である．

異質データ同士のデータマイニング法として正準相関解析法(CCA, canonic al correlation analysis)が知られている[1]. これは 2 つのデータセットにおいて互いの相関を最大にする成分を抽出する方法である．近年では典型的な CCA を改良した手法がいくつか報告されている． 例えば，スパース正準相関解析を用いてパスウェイの遺伝子を予測した研究[2]，カーネル正準相関解析による大腸菌のオペロン予測問題に適用した例[3]など，いずれもプラットフォームの異なる複数のデータセットの相関関係に注目している．特に，正準相関解析にカーネル法を適用した方法[4, 5]では，カーネルの選択によって多変量間の非線形的な関係を特徴空間に射影して線形関係を解析することができた．また正則化項が導入されており高次元データで起こる不良設定問題が解決されている． そのため，高次元かつ複雑な構造を持つ生物データを解析する方法として極めて有効であると考えられる．

一方，J. Lamb ら(2006)は，様々な薬物による影響を比較した遺伝子発現データベースを発表した[6]．この研究では薬の種類，投与量，細胞の種類などの条件の類似関係を比較している．しかし，薬の作用機序を理解し，副作用の原因を把握するためには，条件の比較に加えてそれに関わる詳細な分子メカニズムの同定が必要である．また，作用機序の不明な薬については分子レベルでの潜在的な影響を調べることも要求されている．

そこで本研究では，カーネル正準相関解析法という優れたアプローチによって，抗がん剤の暴露によって活性化される共通パスウェイを検出することを目的とした．

[1] 東京医科歯科大学 大学院生命情報科学教育部
[2] 産業技術総合研究所 生命情報工学研究センター
[3] お茶の水女子大学 生命情報学教育研究センター

[1] School of Biomedical Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
[2] National Institute of Advanced Industrial Science and Technology (AIST), Computational Biology Research Center, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
[3] Ochanomizu University, Center for Informational Biology, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan

Vol.2010-BIO-20 No.9
2010/3/5

2. 正則化正準相関解析

(1) 正準相関解析

正準相関分析とは対象となる 2 組の変数群において，それぞれの合成変数 (＝正準変数)を作り，その合成変量間の相関係数が最大になるような重み係数を求める多変量解析手法である．

ここで，2 つのデータ行列 $\boldsymbol{X}_a \in \Re^{N\times r}$， $\boldsymbol{X}_b \in \Re^{N\times(p-r)}$ があるとし，このデータに対する正準変数を，

$$\boldsymbol{f}_a \equiv (\boldsymbol{X}_a)^T \boldsymbol{w}_a, \quad \boldsymbol{f}_b \equiv (\boldsymbol{X}_b)^T \boldsymbol{w}_b$$

とする．但し， $\boldsymbol{w}_a \in \Re^r$ と $\boldsymbol{w}_b \in \Re^{p-r}$ は正準相関解析のパラメータである．ここで，2 つの合成変量間の相関係数は，

$$\rho(\boldsymbol{f}_a, \boldsymbol{f}_b) = \frac{\langle \boldsymbol{f}_a, \boldsymbol{f}_b \rangle}{\|\boldsymbol{f}_a\| \|\boldsymbol{f}_b\|}$$

と表される．但し， $\langle \boldsymbol{f}_a, \boldsymbol{f}_b \rangle$ は内積を意味する．

この $\rho(\boldsymbol{f}_a, \boldsymbol{f}_b)$ を，

$$\boldsymbol{w}_a^T \boldsymbol{X}_a \boldsymbol{X}_a^T \boldsymbol{w}_a = \boldsymbol{w}_b^T \boldsymbol{X}_b \boldsymbol{X}_b^T \boldsymbol{w}_b = 1$$

のもとで最大にする $\boldsymbol{w}_a \in \Re^r$ と $\boldsymbol{w}_b \in \Re^{p-r}$ を求める．但し，最大の相関係数に対する正準変数を第 1 正準変数，次に大きい相関係数に対する正準変数を第 2 正準変数という．

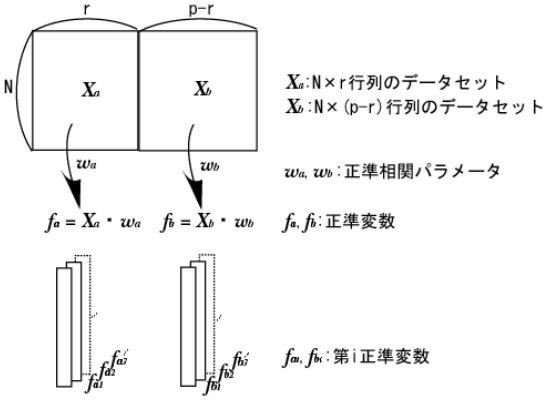

図 1 正準相関解析の概念図

(2) リッジ正準相関解析

リッジ正準相関解析では，2 つの正則化パラメータ γ_a， γ_b を用い，正則化相関係数，

$$\rho(\boldsymbol{f}_a, \boldsymbol{f}_b) = \frac{\langle \boldsymbol{f}_a, \boldsymbol{f}_b \rangle}{\sqrt{\|\boldsymbol{f}_a\|^2 + \gamma_a \boldsymbol{w}_a^T \boldsymbol{w}_a}\sqrt{\|\boldsymbol{f}_b\|^2 + \gamma_b \boldsymbol{w}_b^T \boldsymbol{w}_b}}$$

を最大する $\boldsymbol{w}_a \in \Re^r$ と $\boldsymbol{w}_b \in \Re^{p-r}$ を求める方法である．但し， $\boldsymbol{w}_a^T \boldsymbol{X}_a \boldsymbol{X}_a^T \boldsymbol{w}_a + \gamma_a \boldsymbol{w}_a^T \boldsymbol{w}_a = \boldsymbol{w}_b^T \boldsymbol{X}_b \boldsymbol{X}_b^T \boldsymbol{w}_b + \gamma_b \boldsymbol{w}_b^T \boldsymbol{w} = 1$ とする．正則化パラメータは，所与のデータへの過剰なフィッティングを防ぐ．

(3) カーネル正準相関解析

カーネル正準相関解析は，基本的にリッジ正準相関解析にカーネル法を適用したものである． カーネル法では，ある写像 ϕ_a， ϕ_b が存在すると仮定して，

$$\mathbf{K}_a(\boldsymbol{x}_a, \boldsymbol{x}_a') = \langle \phi_a(\boldsymbol{x}_a), \phi_a(\boldsymbol{x}_a') \rangle, \quad \mathbf{K}_b(\boldsymbol{x}_b, \boldsymbol{x}_b') = \langle \phi_b(\boldsymbol{x}_b), \phi_b(\boldsymbol{x}_b') \rangle$$

を満たすような関数 $\mathbf{K}_a, \mathbf{K}_b$ を用いる． このような関数はカーネル関数と呼ばれる．カーネル関数によって非線形な成分の抽出が可能となる．このカーネル関数を用いると正準変数は，

$$\boldsymbol{f}_a \equiv \phi_a(\boldsymbol{X}_a)^T \boldsymbol{w}_a = \mathbf{K}_a \boldsymbol{\alpha}_a, \quad \boldsymbol{f}_a \equiv \phi_b(\boldsymbol{X}_b)^T \boldsymbol{w}_b = \mathbf{K}_b \boldsymbol{\alpha}_b$$

と表される．但し， $\boldsymbol{\alpha}_a$ および $\boldsymbol{\alpha}_b$ はカーネル正準相関解析のパラメータである． 正則化相関係数は，

$$\rho(\boldsymbol{f}_a, \boldsymbol{f}_b) = \frac{\boldsymbol{\alpha}_a^T \boldsymbol{K}_a \boldsymbol{K}_b \boldsymbol{\alpha}_b}{\sqrt{\boldsymbol{\alpha}_a^T \boldsymbol{K}_a^2 \boldsymbol{\alpha}_a + \gamma_a \boldsymbol{\alpha}_a^T \boldsymbol{\alpha}_a}\sqrt{\boldsymbol{\alpha}_b^T \boldsymbol{K}_b^2 \boldsymbol{\alpha}_b + \gamma_b \boldsymbol{\alpha}_b^T \boldsymbol{\alpha}_b}}$$

と表される．但し， $\boldsymbol{\alpha}_a^T \boldsymbol{K}_a^2 \boldsymbol{\alpha}_a + \gamma_a \boldsymbol{\alpha}_a^T \boldsymbol{\alpha}_a = \boldsymbol{\alpha}_b^T \boldsymbol{K}_b^2 \boldsymbol{\alpha}_b + \gamma_b \boldsymbol{\alpha}_b^T \boldsymbol{\alpha}_b = 1$ であるとする．

正確には，分母の $\boldsymbol{\alpha}_a^T \boldsymbol{\alpha}_a$ は $\boldsymbol{\alpha}_a^T \boldsymbol{K}_a \boldsymbol{\alpha}_a$， $\boldsymbol{\alpha}_b^T \boldsymbol{\alpha}_b$ は $\boldsymbol{\alpha}_b^T \boldsymbol{K}_b \boldsymbol{\alpha}_b$ となければリッジ正準相関解析をカーネル化したものにならない．しかし，そうすると正則化相関係数を最大化するときに特異行列が現れてしまい数値計算との問題が生じる．故に，正則化項として $\boldsymbol{\alpha}_a^T \boldsymbol{\alpha}_a$ および $\boldsymbol{\alpha}_b^T \boldsymbol{\alpha}_b$ に正則化パラメータをかけたものを用いた[5].

3. 実験データ

3.1 遺伝子発現データ

(1) データの収集

撹乱物質をヒト培養細胞に暴露した遺伝子発現データをThe Connectivity Map[6] のデータベースであるhttp://www.broadinstitute.org/cmap/よりダウンロードした．また，その化合物が抗がん剤かどうかの判別は，STITCH [7]のデータベースにある化合物とATC (Anatomical Therapeutic Chemical Classification System)分類の対応表を用いて行った．ATC分類とは，解剖治療化学分類法と呼ばれ医薬品の分類に用いられており，特にL01で始まるコードの医薬品は抗悪性腫瘍の治療に使用されている．具体的な抗がん剤の名前やATCコードおよび薬理学的分類等を付録Aに表した．

◆ 暴露物質：抗がん剤 16 種
◆ 暴露対象：ヒト培養細胞 MCF7(乳癌培養細胞)，PC3(前立腺癌培養細胞)およびHL60(骨髄性白血病)の 3 種類
◆ マイクロアレイの種類：GeneChip® HT Human Genome U133 Array (Affymetrix)
◆ プローブセットの数：22277 個
◆ 薬物処理したサンプル数：66 トリートメント（薬物の種類，濃度，培養細胞の種類による分類）
◆ コントロールサンプルの数：249 コントロール（一つのトリートメントに対して細胞の種類が同一な 1～6 つの溶媒コントロールが存在）
◆ 暴露時間：6 時間

(2) 前処理

遺伝子発現データの前処理は，RMA(Robust Multichip Average)法によって行った[10]. その後，溶媒コントロールに対するトリートメントの発現量の比(=fold-change)を式(1)より求めた．

$$\text{fold-change} = \log_2 (V_t / V_c) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

(V_t: トリートメントの発現値, V_c:コントロールの発現値)

但し，一つのトリートメントに対して複数のコントロールが存在した場合は，コントロール同士の発現量の幾何平均値を用いて比を計算した．fold-change を 66 対のトリートメントおよびコントロールの 22277 個の遺伝子に対してそれぞれ計算し，データセット $\boldsymbol{X}_b$ とした. また, 遺伝子の ID としては Entrez Gene ID を用いた. Entrez Gene ID は 13524 種類あり，複数のプローブセット ID が対応した場合は四分位数範囲(IQR)を計算して，サンプル間においてばらつきの大きいプローブセット ID のみを対応づけした．

3.2 遺伝子パスウェイ

遺伝子のパスウェイ情報として KEGG パスウェイデータを用いた．現在分類されているヒトのパスウェイおよび遺伝子の種類はそれぞれ 197 個, 5185 個であった． これを，あるパスウェイに対してその遺伝子が含まれている場合は 1, そうでない場合は 0 とバイナリデータとして作成し，データセット $\boldsymbol{X}_a$ とした．遺伝子 ID としては 3.1(2) と同様に Entrez Gene ID を用いた．
最終的に， $\boldsymbol{X}_a$ および $\boldsymbol{X}_b$ において共通な Entrez Gene ID は 4287 個あった．

4. 実験

4.1 カーネル正準相関解析

上記の方法によって，作られた $\boldsymbol{X}_a$ および $\boldsymbol{X}_b$ の最終的なデータセットの詳細を表 1 に示した．

表 1 データセットの内容

	$\boldsymbol{X}_a$	$\boldsymbol{X}_b$
データの種類	パスウェイ	遺伝子発現プロファイル
行の情報（行数）	遺伝子(4287)	遺伝子(4287)
列の情報（列数）	パスウェイの ID (197)	トリートメント(66)

4.2 パラメータの調整

(1) 正則化パラメータ

正則化パラメータ γ_a および γ_b の値を決定するために，10- fold クロスバリデーション法を用いた．候補の値として γ_a = 0.01, 0.1, 1, 10, 100, 1000, 10000 および γ_b =0.01, 0.1, 1, 1 0, 100, 1 000, 1 0000 を選び，すべての組み合わせに対して次の手順を実行した．

データセットを 10 グループに分割し，9 グループを訓練用，1 グループを評価用とした．訓練用データから $\boldsymbol{\alpha}_a$ および $\boldsymbol{\alpha}_b$ を求め，そのパラメータを使って得られる評価用データの正準変量間の相関係数を計算した．これをローテーションして１０回繰り返し，10 個の相関係数の平均を算出した．これを（γ_a，γ_b）のすべての値の組み合わせに対して実行しもっとも最大の固有値に対する正準変数(第 1 正準変数) の平均相関係数が高かった γ_a = 0.1，γ_b =1 を正則化パラメータとして採用した．

(2) **RBF カーネルパラメータ**

本研究では，RBF(Radial basis function; 動径基底関数)関数をカーネル関数として用いた．また，この RBF カーネルのパラメータ σ_a および σ_b は， $\boldsymbol{X}_a$ および $\boldsymbol{X}_b$ のそれぞれのデータセットにおける遺伝子間のユークリッド距離を表す距離行列 $\boldsymbol{D}_a \in \boldsymbol{R}^{4287\times4287}$ および $\boldsymbol{D}_a \in \boldsymbol{R}^{4287\times4287}$ を作成し，それらの要素の平均値である $\sigma_a = 2.19$， $\sigma_b = 2.95$ をパラメータとして使用した[9].

5. 結果

5.1 第 1 正準変数と相関の高い変数(r=0.98)

第 1 正準変数に対する評価用データセットの変数同士の相関係数は, 0.98 であった．さらに, 求められた正準変数と最初の変数群(パスウェイ変数=197 種類，トリートメント=66 種類）とのピアソンの積率相関係数(=構造係数, S-score)を計算し，互いの相関係数の傾向と有意にプラスの相関またはマイナスの相関している変数を求めた．
その結果，Ascorbate and alda rate metabolism と Pentose and glucur onate interconversions パスウェイは最も S-score が高く，強い相関関係にあることを示唆した．また，このパスウェイと含め，Porphyrin and chloro phyll m etabolism や Androgen and estro gen metabolism などのパスウェイ変数と, ニトロソ尿素系のアルキル化剤である carmustine, semustine が正の相関関係にあった．一方，Streptozotocin, Altretam ine 等とは負の相関を示した．

表 2　第 1 正準変数と相関の高いパスウェイ変数(S-score ≧ 0.20)

パスウェイ ID†	パスウェイ名	S-score
00053	Ascorbate and aldarate metabolism	-0.84
00040	Pentose and glucuronate interconversions	-0.80
00860	Porphyrin and chlorophyll metabolism	-0.64
00150	Androgen and estrogen metabolism	-0.60
00500	Starch and sucrose metabolism	-0.58
00983	Drug metabolism - other enzymes	-0.58
00830	Retinol metabolism	-0.53
00980	Metabolism of xenobiotics by cytochrome P450	-0.51
00982	Drug metabolism - cytochrome P450	-0.49

†パスウェイ ID は KEGG パスウェイにおけるマップ ID を表す

表 3　第 1 正準変数と相関の高いトリートメント変数(S-score 上位 10 個)

トリートメント ID¶	S-score	化合物名
6888	-0.18	carmustine
7487	-0.16	semustine
7540	-0.15	semustine
6098	0.12	Streptozotocin
4627	0.11	Altretamine
5571	0.10	retinoic acid
6681	0.10	Etoposide
6914	-0.10	carmustine
7089	-0.10	lomustine
7050	-0.095	lomustine

¶トリートメント ID として Connectivity map のデータベースの認識コードを用いた

5.2 第 2 正準変数と相関の高い変数(r=0.44)

第 2 正準変数に対して高い相関を示しているパスウェイ変数としては，細胞周期, DNA 複製などが得られた． これらに対し，正の相関を示したのは Streptozotocin であった，負の相関を示したのは paclitazel, daunorubicin HCl 等であった．

表 4　第 2 正準変数と相関の高いパスウェイ変数(S-score ≧ 0.20)

パスウェイ ID	パスウェイ名	S-score
04110	Cell cycle	0.32
03030	DNA replication	0.26
04080	Neuroactive ligand-receptor interaction	-0.25
04060	Cytokine-cytokine receptor interaction	-0.25
03040	Spliceosome	0.25
00240	Pyrimidine metabolism	0.22
03010	Ribosome	0.20
04120	Ubiquitin mediated proteolysis	0.20
03420	Nucleotide excision repair	0.20

Vol.2010-BIO-20 No.9
2010/3/5

表 5　第 2 正準変数と相関の高いトリートメント変数(S-score 上位 10 個)

トリートメント ID	化合物名	S-score
2535	Streptozotocin	0.30
7193	Streptozotocin	0.20
6720	Paclitaxel	-0.18
7050	lomustine	-0.17
7507	daunorubicin HCl	-0.17
3241	Etoposide	-0.15
5583	nordihydroguaiaretic acid	0.14
1636	retinoic acid	-0.14
5688	Altretamine	-0.13
5320	Paclitaxel	-0.13

6.　考察

まず，第 1 正準変数と相関の高いトリートメント変数を検討したところ，ニトロソ尿素系のアルキル化剤とそれ以外の薬剤との相違が見られた．ニトロソ尿素系のアルキル化剤である carmustine, semustine および lomustine ではビタミン代謝，ホルモン代謝，糖代謝，薬物代謝など生体全般的な代謝関連のパスウェイの促進作用が見られるが，逆にそれ以外の Streptozotocin, Altretamine, retinoic acid および Etoposide では，これらのパスウェイが全体的に抑制されている． これらのパスウェイは，細胞が取り入れた外部の物質を分解し，必要なものを生成するために必須な生体反応である．したがって，今回の解析に用いたニトロソ尿素系のアルキル化剤以外の抗がん剤においてはこれらのパスウェイの抑制が作用機構となっている可能性がある．

次に，第 2 正準変数との相関の高い変数について考察する．それぞれの変数群において，第 1 正準変数と比べて S-score は低いものの，特徴的なパスウェイ群との相関関係が見られた．例えば，細胞周期，DNA 複製，スプライソソーム，ピリミジン代謝，リボソーム，ユビキチン媒介蛋白質分解，ヌクレオチド除去修正は細胞増殖に関わるものであるが，いずれも第 2 正準変数と負に相関している．したがって，これらは薬剤の抗腫瘍活性に関っていると考えられる．一方で，Streptozotocin はこれらのパスウェイと負の相関を示した．Streptozotocin は化合物分類ではニトロソ尿素系であるが，薬理学的分類ではアルキル化剤ではなく抗生物質に含まれる．また，日本ではこの薬剤が主に動物実験だけで使われており，発がん物質として分類されるケースもあるなど非常に細胞毒性が強い物質である．そのため，暴露された細胞が損傷され，細胞周期，DNA 複製などの細胞周期関連パスウェイが過剰に働いているのが，Streptozotocin と正の相関を示した原因として考えられる．

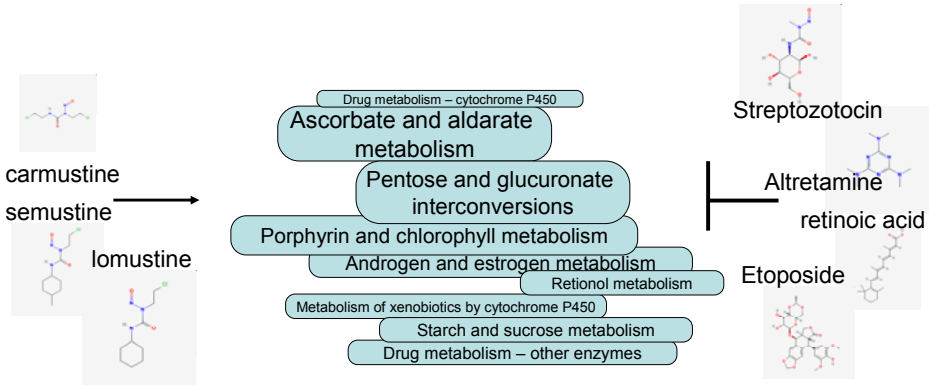

図 2　抗がん剤とパスウェイの相関関係の仮説図 (第 1 正準変数関連)

7.　おわりに

我々は，正則化項およびカーネル法を導入した正準相関解析法を用いて，抗がん剤の影響によって活性または抑制されるパスウェイを同定した．　まだ，正準相関解析は遺伝子発現などのオミックスデータ解析に応用された例が少なく，結果の解釈，性能の比較の方法が定着していない傾向がある．本研究では，解釈において S-score を用い，正準変量と各変数の相関係数を調べることにより，変数同士がどのように正または負の相関関係とその度合いについて意味づけすることが出来た．今後の課題として，さらにカーネル関数の種類を検討し，それによって抽出される正準変量の特徴を比較していく方針である．

謝辞　薬物関連のデータベースにおいて助言をいただいた東京医科歯科大学の高井貴子先生，その他ご協力頂いた皆様に，心から感謝の意を表します．

参考文献

1) W.T. Anderson, An introduction to multivariate statistical analysis, John Wiley & Sons. 1984.
2) S. Waaijenborg and A.H.Z., *Sparce canonical correlation analysis for identifitying, connecting and completing gene-expression networks.* BMC Bioinformatics, 2009. **10**(315).
3) Y. Yamanishi, J.V., A. Nakaya and M. Kanehisa, *Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis.* Bioinformatics, 2003. **19**.
4) S. Akaho, *A kernel method for canonical correlation analysis.* Workshop on information-based induction sciences (IBIS2000), 2000.
5) Kuss, M. and T. Graepel, *The Geometry of kernel canonical correlation analysis.* 2003.
6) J. Lamb et al., *The Connectivity Map: a new tool for biomedical research.* Nat Rev Cancer, 2007. **7**(1): p. 54-60.
7) M. Kuhn, D.S., A. Franceschini, M. Campillos, C.V. Mering, L.J. Jensen, A.B. and P. Bork, *STITCH 2: an interaction network database for small molecules and proteins.* Nucleic Acids Research, 2009.
8) NCBI PUBCHEM;http://pubchem.ncbi.nlm.nih.gov/
9) T. Kato, H. Kashima., M. Sugiyama and K. Asai, *Conic Programming for Multi-Task Learning.* IEEE Transactions on Knowledge and Data Enginieering, Accepted.
10) R. A. Irizarry, B.H., F. Collin, Y. D. Beazer-Barclay, K.J. Antonellis, U. Scherf and T. P. Speed, *Exploration, normalization, and summaries of high density oligonucleotide array probe level data* Biostatistics, 2003, **4**(2): p. 249-264.

付録

付録 A 薬の名前と CID，ATC コード，病理学的および化合物的分類

番号	化合物名	ATC 分類コード [7]	病理学的分類[8]	化合物分類[8]
1	Altretamine	L01XX03	抗がん剤, アルキル化剤	複素環式化合物
2	carmustine	L01AD01	抗がん剤, アルキル化剤	ニトロソ尿素
3	Chlorambucil	L01AA02	抗がん剤, アルキル化剤	ナイトロジェンマスタード
4	Dacarbazine	L01AX04	抗がん剤, アルキル化剤	複素環式化合物
5	daunorubicin HCl	L01DB02	抗がん剤，抗生物質	多環式炭化水素，芳香族
6	Etoposide	L01CB01	抗がん剤，植物性	多環式炭化水素，芳香族
7	Ifosfamide	L01AA06	抗がん剤, アルキル化剤	ナイトロジェンマスタード
8	Isotretinoin	L01XX22	抗がん剤，膚科用薬物	環状炭化水素
9	lomustine	L01AD02	抗がん剤, アルキル化剤	ニトロソ尿素
10	Methotrexate	L01BA01	代謝拮抗薬，抗がん剤	複素環式化合物
11	Paclitaxel	L01CD01	抗がん剤，植物性	環状炭化水素
12	retinoic acid	L01XX22	抗がん剤	環状炭化水素
13	semustine	L01AD03	抗がん剤, アルキル化剤	ニトロソ尿素
14	Streptozotocin	L01AD04	抗がん剤，抗生物質	ニトロソ尿素
15	vinblastine sulfate	L01CA01	抗がん剤，植物性	複素環式化合物
16	nordihydroguaiaretic acid	L01XX10	抗酸化剤	環状炭化水素