H-019

部分バンドパターンを単位とした染色体領域の抽出・同定*

Chromosome Region Extraction and Classification Based on Local Band Patterns

濱田 智栄子† 山口 傑† 島田 裕充‡ 浅野 克敏‡ 酒井 栄一‡ 大庭 信之§ 佐藤 俊哉§ 阿部 亨¶ 木下 哲男¶

Chieko Hamada† Suguru Yamaguchi† Hiromitsu Shimada‡ Katsutoshi Asano‡
Eiichi Sakai‡ Nobuyuki Ohba§ Toshiya Satoh§ Toru Abe¶ Tetsuo Kinoshita¶

1. はじめに

染色体の光学顕微鏡画像に基づき，種々の異常を視覚的に診断する染色体画像解析は，染色体異常に起因する疾病の予防や治療，染色体異常を誘発する物質の判定試験などにおいて重要な役割を果たす．近年，染色体の視覚的診断に関する知見が深まり，染色体画像解析の重要性が高まるにつれ，染色体画像解析に必須の処理である染色体領域の抽出・同定の自動化が以前にも増して強く求められている．染色体領域を自動的に抽出・同定する手法については多くの研究があり，一部は製品化も行われている．しかし，十分な精度が得られる手法は未だ実現されておらず，実際の染色体画像解析に際しては，専門家による手作業で染色体領域の抽出・同定が行われる場合が多い．

本稿では，染色体領域の高精度かつ柔軟な抽出・同定を実現するために，1 本の染色体領域を複数の部分領域（部分バンドパターン）の連なりと見なし，部分領域単位で探索（抽出・同定）を行う手法を提案する．さらに，効果的探索を実現するための部分バンドパターン決定法を提案し，シミュレーション画像を用いた実験により，その有効性の検証を行う．

2. 染色体領域の抽出・同定

2.1 染色体画像解析の一般的手順

染色体は，遺伝情報を担う細胞核内の生体物質であり，ヒトの正常な細胞核 1 個には，22 種類 2 本ずつからなる 44 本の常染色体と 2 本の性染色体（女性は X 染色体 2 本，男性は X 染色体と Y 染色体 1 本ずつ）の合計 46 本が含まれる．各染色体は，動原体と呼ばれる箇所でくびれており，ここで 2 つの部分に分けられる [1].

染色体の光学顕微鏡画像に基づき，染色体の異常を視覚的に診断する場合，一般に，以下の手順で解析が行われる [2].

1. 染色体を染色し光学顕微鏡画像を撮影
2. 光学顕微鏡画像から染色体領域を抽出
3. 各染色体領域が何番の染色体に対応するかを同定
4. 各染色体領域の状態から種々の異常を視覚的に診断

染色体画像解析には，染色により各染色体に固有の縞（バンド）を染め出し，その状態を撮影した画像を用いる．通常，まず，画像から染色体領域を抽出し，次に，抽出された各領域の特徴を基に，各々が何番の染色体に対応するか同定を行う（図 1）．染色体領域の長さや幅は，同じ種類（番号）の染色体であっても，細胞周期における時期の違い等により異なる．一方，染色体領域同士の長さの比や，各染色体領域での動原体の相対的位置，各染色体領域におけるバンドの現れ方（バンドパターン）などは，ほぼ一定に保たれるため，染色体領域の同定には，これらの特徴が用いられる．同定された結果に基づき，染色体の本数が通常とは異なる数的異常の診断が行われる．さらに，各染色体領域の状態から，遺伝物質の欠失（一部のバンドが欠ける）や重複（余分なバンドが生じる），転座（バンドの位置が通常と異なる）など，染色体の構造異常の診断が可能となる．

2.2 染色体領域抽出・同定のための従来手法

染色体領域を同定する従来手法の多くは，独立の前処理として染色体領域の抽出を行っており，各領域が事前に正しく抽出されていることを前提としている．しかし，染色体領域には，背景と区別し難い箇所や領域の接触・重なりが生じている箇所が存在し，個々の染色体領域の正確な抽出は容易ではない．

染色体領域の同定手法は，各染色体領域の全体的特徴に基づくものと，部分的特徴に基づくものとに大別される [3]. 全体的特徴に基づく手法では，染色体毎に 1 本分全体のバンドパターンを参照パターンとして用意し，抽出された領域のバンドパターンと参照パターンとを比較することで同定を行う [4, 5]. 従って，染色体領域が正確に抽出されても，領域同士が重なった場合や染色体に構造異常が生じている場合など，通常とは異なるバンドパターンが一部に存在する状況では正確な同定が困難となる．一方，部分的特徴に基づく手法では，特徴的な一部のバンドなどの部分的な特徴を用い同定を行う [6, 7]. このため，通常とは異なるバンドパターンが一部に存在する状況へも，ある程度対応可能となる．しかし，全体的特徴を用いる手法に比べ，その同定精度は低いと報告されている [5]. これは，部分的特徴の安定した抽出が難しく，さらに，同定に利用できる情報が少ないためと考えられる．

3. 部分バンドパターンを単位とした染色体領域の抽出・同定

本稿では，従来手法における問題の解決を図り，染色体領域の高精度かつ柔軟な抽出・同定を実現するために，染色体領域の抽出・同定を部分バンドパターン単位で行う手法を提案する．

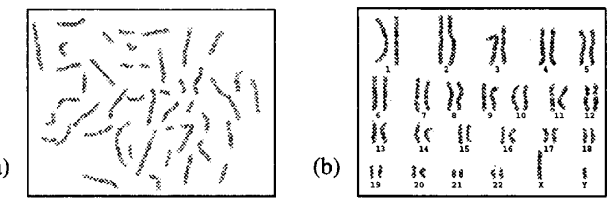

図 1 (a) 染色体の光学顕微鏡画像，(b) 染色体領域の抽出・同定結果

† 東北大学大学院情報科学研究科
‡ （株）日本遺伝子研究所
§ 日本アイ・ビー・エム（株）
¶ 東北大学情報シナジー機構

* 本研究の一部は，文部科学省 科学研究費補助金 萌芽研究，及び，科学技術振興機構 平成 18 年度シーズ発掘試験により行われた．

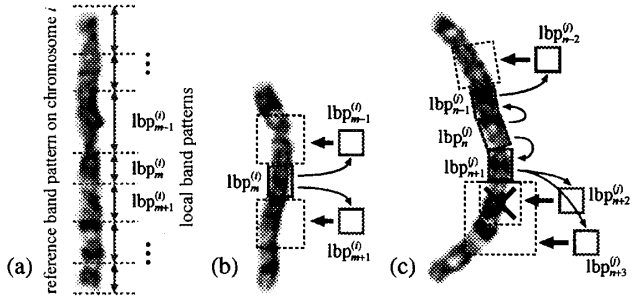

図 2　部分バンドパターンを単位とした染色体領域の抽出・同定

提案手法は，抽出・同定を行う単位として，染色体の番号毎に用意した参照パターンを各々複数の部分に分割したものを用いる（図 2 (a)）．以下では，これを部分バンドパターンと呼び，i 番染色体の m 番部分バンドパターンを $\mathrm{lbp}_m^{(i)}$ で表す．

まず，提案手法では，部分バンドパターンをいくつか選択し，図 2 (b) に示すように，これらに類似する箇所を画像中で探索する．部分バンドパターン $\mathrm{lbp}_m^{(i)}$ に対応する箇所を検出した場合，次に，その近傍で，隣接する部分バンドパターン $\mathrm{lbp}_{m-1}^{(i)}, \mathrm{lbp}_{m+1}^{(i)}$ の対応箇所を探索する．この反復により，最初に検出した箇所を起点に，隣接部分バンドパターンに対応する箇所を順に探索し，1 本の染色体領域全体を芋蔓式に抽出・同定する．さらに，異なる箇所を起点に探索を繰り返すことで，全染色体領域の抽出・同定を行う．

一連の過程で，ある部分バンドパターンの探索に失敗した場合，その箇所では，領域の重なりや染色体の構造異常により，通常とは異なるバンドパターンが生じている可能性が高いと考えられる．そこで，提案手法では，状況に応じて次の探索範囲（画像中の領域）や探索対象（部分バンドパターン）を変更し，このような箇所への対応を図る．例えば，図 2 (c) に示すように，$\mathrm{lbp}_{n+1}^{(j)}$ の近傍で $\mathrm{lbp}_{n+2}^{(j)}$ の探索に失敗した場合，探索範囲を拡大したり探索対象を $\mathrm{lbp}_{n+3}^{(j)}$ に変更する等の制御を行う．

以上のアプローチにより，提案手法では以下が期待できる．

- 部分バンドパターン単位の探索により抽出・同定を同時に行い，さらに，探索結果を相互に利用することで，様々な状態にある染色体領域の高精度・効率的な抽出を実現．
- 探索範囲・対象を状況に応じて制御し，通常とは異なるバンドパターンを示す箇所を除きながら領域の特徴を統合することで，染色体領域の高精度・柔軟な同定を実現．

4. 探索に適した部分バンドパターンの決定法

前章で述べた染色体領域の抽出・同定手法では，探索単位として用いる部分バンドパターンの選び方により効果が大きく変化する．例えば，長い部分バンドパターンを探索単位とすれば，染色体領域の屈曲や，通常とは異なるバンドパターンが生じた箇所への対応が困難となる．一方，短い部分バンドパターンを探索単位とすれば，対応する箇所を画像中で特定し難くなる．本章では，各部分バンドパターンを用いた場合に生じる探索誤りの頻度を推定することで，予め用意した部分バンドパターンの中から探索に適したものを選択する手法を提案する．

いま，部分バンドパターン lbp をテンプレートとしたマッチングにより，lbp に対応する画像中の箇所を探索する場合を考える．画像中の各箇所と lbp との輝度値の差の二乗平均（MSE）に基づき対応箇所を決定するものとし，画像中の箇所 p での MSE を $m(p)$，正解箇所 p_c での MSE を $m(p_c) = m_c$ とすると，$m(p) \leq m_c$ となる p は，lbp に対応すると誤って判定されることになる．$m(p)$ の分布は，対象画像を lbp で走査すれば求めることができる．従って，m_c を推定すれば，lbp に対応すると誤判定される p の総数 N の推定が可能となる．

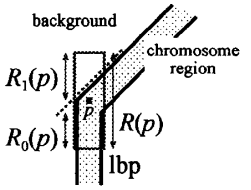

図 3　染色体領域と lbp との関係

染色体領域と lbp との関係を図 3 のようにモデル化し，lbp と重なる画像中の領域を $R(p)$ で表す．$R(p)$ 中で染色体領域と背景に対応する部分を各々 $R_0(p), R_1(p)$ とし，各々の面積比を $1-\alpha(p) : \alpha(p)$，各々での MSE を $m_0(p), m_1(p)$ とすると，$R(p)$ での MSE $m(p)$ は式 (1) で表される．

$$m(p) = (1-\alpha(p)) \cdot m_0(p) + \alpha(p) \cdot m_1(p) \qquad (1)$$

正解箇所 $p = p_c$ に lbp が位置する場合，$m_0(p_c) \approx 0$ と考えられるため $m_c \approx \alpha(p_c) \cdot m_1(p_c)$ となる．p_c は，画像中で様々な状態をとるため，個々の場合について $\alpha(p_c), m_1(p_c)$ を事前に求めることは困難である．しかし，画像中の各箇所での状態から $\alpha(p_c), m_1(p_c)$ の平均 $\overline{\alpha}, \overline{m_1}$ を決定すれば，m_c の平均的な値 $\overline{m_c}$ は式 (2) で推定可能と考えられる．

$$\overline{m_c} \approx \overline{\alpha} \cdot \overline{m_1} \qquad (2)$$

$m(p)$ の平均 $\overline{m}$ は，対象画像を lbp で走査することで，$m_0(p)$ の平均 $\overline{m_0}$ は，lbp を作成するために用意した参照パターンを全て走査することで求めることができる．また，画像中での背景の輝度を一定とみなせば $\overline{m_1}$ も決定できる．これらを用いれば式 (3) で $\overline{\alpha}$ を得ることができ，式 (2) から $\overline{m_c}$ が推定できる．

$$\overline{\alpha} = (\overline{m} - \overline{m_0}) \,/\, (\overline{m_1} - \overline{m_0}) \qquad (3)$$

予め用意した lbp の各々について，対象画像へ適用した場合の MSE の分布および $\overline{m_c}$ を求め，各 lbp を用いた場合に生じる探索誤りの平均的な頻度 $\overline{N}$ を推定すれば，推定される $\overline{N}$ が小さい lbp を探索に用いることで，効果的な染色体領域の抽出・同定が実現できると考えられる．

5. 実験

4 で提案した，探索誤り頻度の推定方法の有効性を検証するために，シミュレーション画像を対象とした実験を行った．

5.1 シミュレーション画像，部分バンドパターン

実験に用いたシミュレーション画像は，ISCN (An International System for Human Cytogenetic Nomenclature) [2] による染色体の模式図（二値で表現した標準的なバンドパターン）を

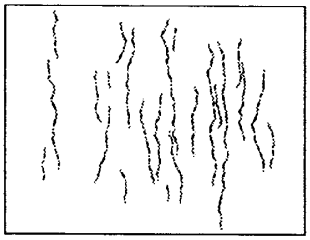

図 4　シミュレーション画像の例

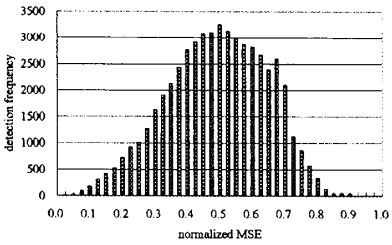

図 5　MSE の分布の例

ランダムに分割・回転したもの（1, . . . , 22, X, Y 染色体の計 24 本）を染色体領域として配置し作成した．なお，シミュレーション画像は 50 枚作成し，各画像のサイズは 1280 × 960 画素とした．シミュレーション画像の例を図 4 に示す．

探索に用いる部分バンドパターンは，模式図中のバンドの境界上に始点・終点を設定し，バンドを取り出すことで作成した．始点・終点の可能な全ての組み合せに基づき作成した部分バンドパターンから，濃淡のパターンが他と重複するものを除き，長さの短いものを 1000 個を選択し実験に用いた．

5.2　実験結果

部分バンドパターンで画像を走査した場合の MSE の分布の例を図 5 に示す．図 5 の横軸は，画像中の背景での値が 1 となるよう正規化された MSE を表し，縦軸は，1 個の部分バンドパターンを 50 枚の画像に適用した場合の MSE の累積出現頻度である．ただし，図 5 のグラフでは，背景での MSE（値が 1 となるもの）を除いている．この結果から分かるように，MSE の分布は正規分布に従うと見なせるため，以下の実験では，探索誤りの平均的な頻度 $\overline{N}$ を式 (4) で推定した．

$$\overline{N} = \overline{A} \int_0^{\overline{m_c}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\overline{m})^2}{2\sigma^2}} dx \qquad (4)$$

ここで，$\overline{A}$ は，画像中で部分バンドパターンが染色体領域と重なる p の総数を，σ^2 は，画像中での MSE の分散を表す．

シミュレーション画像 50 枚に，部分バンドパターン 1000 個を適用し，$\overline{N}$ の実測値（縦軸）と式 (4) による推定値（横軸）を部分バンドパターン毎にプロットした結果を図 6, 7 に示す．図 6 の結果では，$\overline{m_c}$ の実測値を式 (4) に適用し $\overline{N}$ の推定値を求めており，$\overline{N}$ の実測値との相関は $r = .933$ となった．図 7 では，$\overline{m_c} = \overline{\alpha} \cdot \overline{m_1}$ で得られる推定値を式 (4) に適用し $\overline{N}$ を求めており，$\overline{N}$ の実測値との相関は $r = .766$ となった．

以上の結果から確認できるように，各部分バンドパターンを用いた場合の $\overline{N}$ は，正解箇所が未知であっても，提案方法を用いることである程度推定でき，抽出・同定に適した部分バンドパターンを $\overline{N}$ に基づいて事前に選択可能と考えられる．

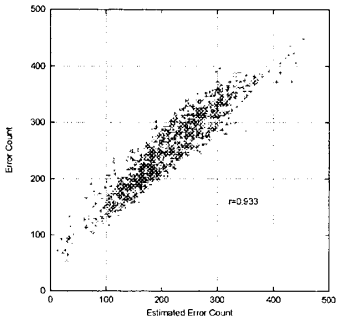

図 6　$\overline{N}$ の推定結果（$\overline{m_c}$ に実測値を用いた場合）

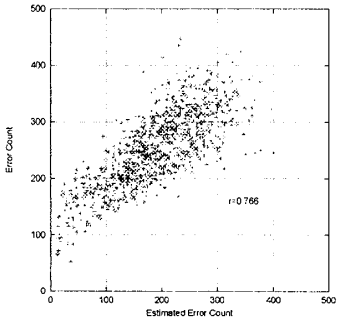

図 7　$\overline{N}$ の推定結果（$\overline{m_c}$ に推定値を用いた場合）

6.　おわりに

本稿では，染色体領域の高精度かつ柔軟な抽出・同定を実現するために，1 本の染色体領域を複数の部分バンドパターンの連なりと見なし，部分バンドパターン単位で探索（抽出・同定）を行う手法を提案した．さらに，効果的探索を実現するための部分バンドパターンの決定法を提案し，シミュレーション画像を用いた実験により，その有効性を示した．

提案する領域抽出・同定手法を実際の染色体画像に適用するためには，画像中で様々な状態をとる染色体領域にも対応できるようにする必要がある．今後，種々の要因が探索の精度や頑健性に与える影響について検証し，染色体領域のより実用的な抽出・同定手法を実現するための検討を進める予定である．

参考文献

[1] 新川 詔夫, 阿部 京子, 遺伝医学への招待 改訂第 3 版, 南江堂, 2005.
[2] ISCN1985: An international system for human cytogenetic nomenclature (1985), S. Karger AG, 1985.
[3] J. Graham and J. Piper, "Automatic karyotype analysis," Meth. Mol. Biol., vol.29, pp.141–185, 1994.
[4] J. Piper and E. Granum, "On fully automatic feature measurement for banded chromosome classification," Cytometry, vol.10, pp.242–255, 1989.
[5] Q. Wu, et al., "Subspace-based prototyping and classification of chromosome images," IEEE Trans. Pattern Anal. Machine Intell., vol.14, no.9, pp.1277–1287, 2005.
[6] F. C. A. Groen, et al., "Human chromosome classification based on local band descriptors," Pattern Recognit. Lett., vol.9, no.3, pp.211–222, 1989.
[7] M. Moradi and S. K. Setarehdan, "New features for automatic classification of human chromosomes: A feasibility study," Pattern Recognit. Lett., vol.27, no.1, pp.19–28, 2006.