

rG4 マイクロアレイ – in vivo rG4 プロファイリングのゴールドスタンダード



RNA G-四重連鎖 (rG4) を in vivo で同定・定量することは、細胞生物学やヒト疾患における rG4 研究する上で不可欠なステップです。次世代シーケンシング (NGS) に基づく技術、例えば G4RP-seq [1]、BG4 uvRIP-seq および DMS-seq Plus RT Stop profiling [2] などは、トランスクリプトーム全体にわたる in vivo rG4 ランドスケープの調査や、さまざまな条件下での定量的変化の評価に用いられてきました。しかし、これらの手法はいずれも正確な rG4 定量には限界があります。

rG4 アレイと G4RP-seq および BG4 uvRIP-seq の比較

G4RP-seq および uvRIP-seq では、一過性に形成される細胞内 G4 構造を保持するためにサンプルを架橋処理が行われます。しかし、間接的な rG4-RNA 相互作用も架橋されてしまう可能性があり、偽陽性結果につながる恐れがあります[3]。

さらに、rG4 捕捉の過程で、BG4 抗体またはリガンドの rG4 への結合は、G3BP1 やヌクレオリンなどの rG4 結合タンパク質によって阻害される可能性があります。これらのタンパク質は生体内で rG4 構造と優先的に相互作用し、結合部位を占有します。例えば、G3BP1 は C 末端 RGG ドメインを介して rG4 に直接結合し、mRNA の安定性を調節することが知られています[4]。これらのタンパク質が rG4 に結合した状態で架橋されると、その後の BG4 抗体や低分子リガンドの結合が阻害され、偽陰性の検出結果が生じる可能性があります。

rG4 アレイでは、架橋の代わりに低分子化学物質である DMS によるメチル化マスキングを用いることで、上記の問題を回避しています。DMS は細胞内および細胞小器官全体に容易かつ迅速に浸透します。DMS メチル化から保護された in vivo rG4 領域が、結合タンパク質を一切含まない精製 RNA において、in vivo で再フォールディングされ、G4 抗体による捕捉および検出を可能にします。Arraystar rG4 マイクロアレイプロファイリングでは、細胞溶解前に in vivo で DMS 処理を行うことでネイティブ rG4 を標識し、その後、抗体濃縮のために in vitro で rG4 を再フォールディングすることで、生体内 rG4 を検出・定量します。これにより、トランスクリプトーム全体におけるネイティブな in vivo rG4 状態の定量精度を大幅に向上させます。

rG4 アレイと DMS-seq + RT Stop Profiling の比較

DMS-seq + RT Stop Profiling 法は、in vivo での DMS 処理ステップに続いて RT Stop プロファイリングを行う手法です。折り畳まれていない状態の rG4 領域のグアニン残基はその N7 位で DMS によりメチル化されますが、折り畳まれた rG4 構造のグアニンはメチル化されません[2]。In vivo の rG4 構造によって DMS メチル化から保護されたグアニンは in vitro で rG4 に再フォールディングできませんが、メチル化されたグアニンにはできないため、in vivo における rG4 の折り畳み状態を間接的に推定することが可能になります。

しかし、DMS はアデノシン (A) の N1 位やシトシン (C) の N3 位もメチル化する可能性があります (図 1)。DMS-seq と RT Stop プロファイリングを組み合わせた場合、逆転写酵素は折り畳まれた rG4 に遭遇した時だけでなく、m1A または m3C に遭遇した時にも停止し、鎖の伸長が停止します。その結果、RT は折り畳まれた rG4 部位だけでなく m1A または m3C 部位でも停止する可能性があり、真の rG4 部位の定量が不正確になる恐れがあります。

Arraystar rG4 マイクロアレイプロファイリングでは、G4 特異的抗体 BG4 を用いて rG4 構造を含む RNA 断片をアフィニティー補足します。これにより、逆転写酵素の停止に依存する場合よりもはるかに信頼性の高い rG4 シグナルが得られます。補足された RNA は、AlkB 処理によってメチル基が除去され、阻害なくバイアスのない逆転写反応を経て cRNA が作製され、rG4 マイクロアレイ検出が行われます。このように、in vivo での DMS 処理と in vitro での再フォールディングにより、真のネイティブ rG4 構造が保持・回復され、さらに抗 G4 抗体 BG4 を用いることで、真の rG4 含有 RNA 断片を効率的かつ特異的に捕捉できます。最後に、rG4-RNA は Arraystar rG4 マイクロアレイを用いて検出され、シーケンス解析よりもはるかに高感度です。総合的に見て、Arraystar rG4 アレイは、in vivo rG4 プロファイリングにおいて最も正確かつ高感度なプラットフォームと言えます。

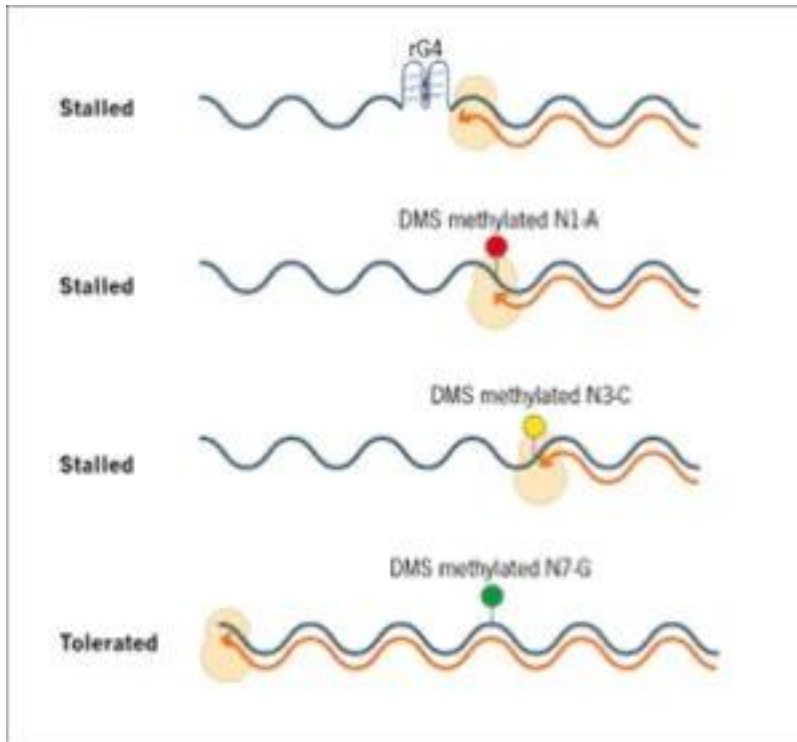


図 1. DMS-seq + RT Stop Profiling

逆転写酵素 (RT) は、折り畳まれた rG4 に遭遇した時だけでなく、DMS メチル化された m1A または m3C に遭遇した時にも停止し、cDNA 鎖の合成が停止するため、rG4 部位の誤同定につながる。

RNA シーケンスは、低存在量の rG4 の定量には不向き

生細胞 (in vivo) における rG4 の形成は、rG4 研究において重要な情報です。主な課題の一つは、rG4 の折り畳みと展開が生体内では急速かつ動的に起こり、自然な平衡状態が非フォールド状態に偏っていることです。このため、細胞内の rG4 含有 RNA の存在量は非常に低いです。

RNA-seq では、シーケンスリードの大部分は高存在量の RNA (例えばハウスキープ遺伝子) に由来する一方、低存在量の RNA はカバレッジが低くなります。リード数が少ないと、RNA 断片の検出感度が低下し、定量の信頼性も下がります。カバレッジ深度を増やしても、低存在量の転写産物の感度と精度は直線的に向上するわけではなく、深い RNA-seq カバレッジにおいても、その効果はすぐに頭打ちとなってしまう。そのため、もともと低存在量である rG4-RNA は、RNA-seq に基づく手法では検出が困難であり、定量化も不十分となります。

例えば、rG4 形成についてよく研究されている RNA である lncRNA-MALAT1 は、シーケンスによる RNA G-四重鎖のグローバルマッピング (G4RP-Seq; Global mapping of RNA G-quadruplexes by sequencin) [3]において、この制限に直面しています。つまり、G4RP-Seq では、MALAT1 からの有意な rG4 シグナルを検出するために、G4 を人工的に誘導および安定化できる G4 安定化リガンド BRACO-19 および RHPS4 を追加する必要があります[1]。しかし、これは通常の生理的条件下での自然で一過性の rG4 を観察するには適していません。

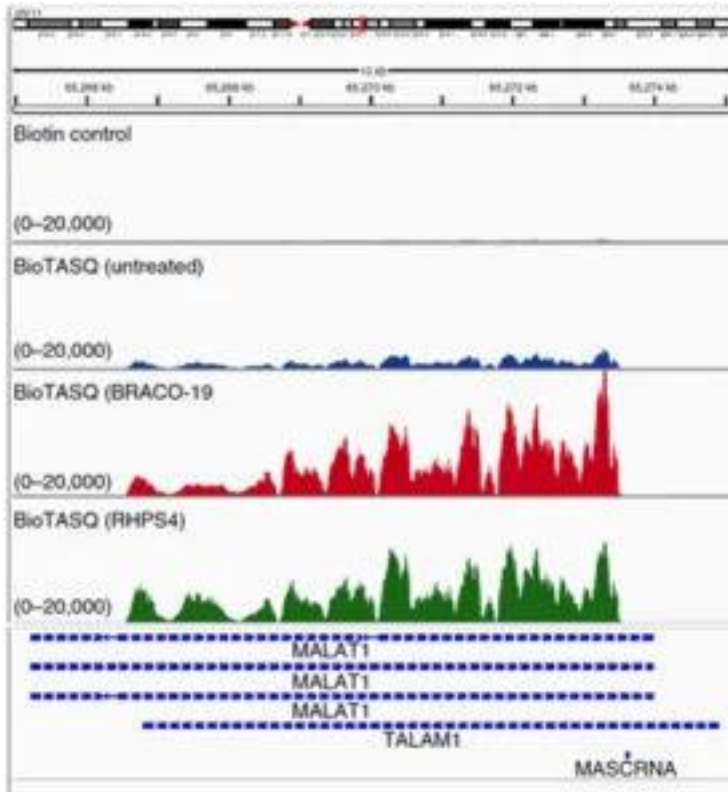


図 2. ビオチンコントロール、未処理、BRACO-19 処理、および RHPS4 処理サンプルの G4RP-Seq からの MALAT1 rG4 シグナル[3] G4 安定化リガンド BRACO-19 または RHPS4 の処理は高い rG4 シグナルを生成するが、rG4 は人工的に誘導され安定化された G4 であり[1]、生体内のネイティブな一過性 rG4 状態を表すものではなく、通常の生理的条件下でのネイティブな rG4 を観察するのに理想的ではない。

さらに、RNA シークエンシング (RNA-seq) では、ライブラリーサイズと遺伝子長の差を正規化するために RPKM (Reads Per Kilobase of transcript per Million reads mapped) が用いられます。しかし、RPKM 値はあるサンプル中の転写産物の相対的な存在量を示すものであり、その量を示すものではありません。RNA 組成の変動が RNA プロファイル全体に影響を与えるため、RPKM を用いて異なるサンプル間で転写産物の発現レベルを比較すると誤解を招く可能性があります。これらの違いは、サンプル間比較を行う際に RNA 発現レベルを歪める可能性があります。したがって、RNA-seq による rG4 プロファイリングは、異なるサンプル間で rG4 の存在量を比較するのに適していません。

rG4 マイクロアレイプロファイリングでは、RNA ターゲットは、たとえ高存在量であっても他の RNA 配列とは独立して、その配列特異的なプローブとハイブリダイズされます。RNA シークエンスでは、高存在量の RNA (例えば、ハウスキーピング遺伝子 RNA) が低存在量の RNA の検出範囲を圧迫する可能性があります。マイクロアレイでは低存在量の転写産物の検出にはほとんど影響を与えません。したがって、マイクロアレイを用いたプロファイリングは、発現レベルが低いという特徴を持つ rG4-RNA を定量化するのに特に適しています。

References

1. Yang SY et al: Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat Commun* 2018, 9(1):4730.[PMID: 30413703]
2. Guo JU, Bartel DP: RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* 2016, 353(6306).[PMID: 27708011]
3. Yang SY, Monchaud D, Wong JMY: Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq. *Nat Protoc* 2022, 17(3):870-889.[PMID: 35140410]
4. He X, Yuan J, Wang Y: G3BP1 binds to guanine quadruplexes in mRNAs to modulate their stabilities. *Nucleic Acids Res* 2021, 49(19):11323-11336.[PMID: 34614161]
5. Jiang L et al: Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 2011, 21(9):1543-1551.[PMID: 21816910]
6. Labaj PP et al: Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics* 2011, 27(13):i383-391.[PMID: 21685096]
7. Toung JM, Morley M, Li M, Cheung VG: RNA-sequence analysis of human B-cells. *Genome Res* 2011, 21(6):991-998.[PMID: 21536721]
8. Kretz M et al: Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* 2012, 26(4):338-343.[PMID: 22302877]
9. Xu W et al: Human transcriptome array for high-throughput clinical studies. *Proc Natl Acad Sci U S A* 2011, 108(9):3707-3712.[PMID: 21317363]