

原來擺對位置很重要-淺談人類免疫缺陷病毒的基因表現

陳珩昌

台灣大學醫學院微生物研究所
Epigenetics of Infectious Diseases Research Group
ŁUKASIEWICZ – PORT, Wrocław, POLAND

壹、從果蠅的位置效應花斑看潛伏在人類染色體中的人類免疫缺陷病毒基因表現

緣起：2011 初我離開柏林到巴賽隆納的一所研究中心（The Centre for Genomic Regulation¹）開始我在地中海的新生活。在那裡我遇到了一位不論在當時或是現在對我研究的指導與啟發都極為重要的一位科學家 — Filion 博士²。他最為重要的學術著作之一即是利用 DamID 技術^{1,3}將果蠅染色體區分為五區²；每區的染色體因為有不同的 DNA 結合蛋白參與作用，造成在各區的基因，其基因表現並不相同。

在果蠅生物學中有一項重要的發現即是位置效應花斑（position effect variegation, PEV）^{3,4}。在一般的情況下，果蠅白眼基因的表現造成果蠅在外觀上的眼色為紅色。科學家米勒發現，如果將同一個果蠅白眼基因移置到靠近染色體著絲粒（centromere）的位置，果蠅的眼色則成為紅白相間。換句話說，位置效應花斑即是指同一個基因在染色體不同的位置會造成基因不同的表現。

你現在可能會很納悶為什麼我們會先

從調控果蠅基因表現的機轉說起。就研究基因表現調控的層面而言，人類免疫缺陷病毒（Human Immunodeficiency Virus type I, HIV-1）其實是除了果蠅之外，另一個可以很好用來研究位置效應花斑對基因表現的影響的天然生物模型。在人類免疫缺陷病毒感染宿主細胞的過程中，有一項重要的步驟是將病毒經反轉錄而形成的互補脫氧核糖核酸（cDNA）插入宿主的染色體中⁴（為了方便閱讀，文後將病毒的互補脫氧核糖核酸統稱為「病毒 DNA」）。在當時已經有許多的研究證實病毒 DNA 是有選擇性地插入宿主的染色體⁵⁻⁷。然而，這種非隨機性插入宿主染色體的現象，對於病毒的致病機轉，例如病毒進入潛伏態（latency）⁵與否，及何時進入潛伏態在當時（2012 年到 2013 年）還是知道的很有限。

貳、人類免疫缺陷病毒 DNA 插入染色體的位置影響病毒基因表現

如果我們將每一隻感染細胞的病毒視為一段相同的基因序列（事實上每隻病毒

的基因序列因為反轉錄的過程以及其它種種原因並不是全然相同)，那麼插入在染色體不同位置的原病毒就可以想成是將調控果蠅眼色的果蠅白眼基因放置於染色體的不同位置。如此一來，我們所要探討的問題就變得很單純：是否人類免疫缺陷病毒的基因表現會因為原病毒插入染色體位置的不同而有所改變。我們的假設很簡單：人類免疫缺陷病毒的基因表現是可以被類似位置效應花班的機轉所調控（圖一）。如果這個假設成立的話，染色體上的哪些區域對於原病毒基因表現是重要的呢？

在當時已經有一些零星的研究^{8,9}顯示原病毒的基因表現，在染色體不同的位置會有所不同。然而受到當時技術上的限制，我們並無法大量地測量每隻原病毒的基因表現。為了能夠同時測量每隻原病毒的基因表現並且連結每隻原病毒在染色體上相對應的插入位置，Filion 博士和我決定建立一個分子生物技術來解決這個問題。我們稱這個技術為 Barcode HIV ensembles (B-HIVE)^{10,11}。這個技術的前身是 Thousands of Reporters Integrated in Parallel (TRIP)^{12,6}；而我們現在則是將此技術的原理直接應用在天然的載體（vector）— 病毒上^{10,11}。簡單地說，我們將一段由二十個隨機挑選的核苷酸所組成的 DNA 序列（我們稱此段 DNA 序列為「DNA 條碼（DNA barcode）」）放入每隻病毒的基因體中（圖二）；理論上，每隻病毒都會帶有不同的 DNA 條碼，在病毒感染宿主細胞後，我們就可以透過追蹤每個 DNA 條碼所在染色體

的位置以及測量每個條碼 RNA 的表現量，來描繪出每隻原病毒 DNA 插入染色體的位置及其相對應的原病毒基因表現¹⁰（圖二）。再結合目前高通量定序（high-throughput sequencing）⁷的技術，我們也就可以在短時間內得到大量的數據來進行在統計上有意義的分析。

參、DNA 上的強化子對於人類免疫缺陷病毒基因表現的影響

利用 B-HIVE 技術，我們得到了一些有趣而且重要的結果。首先，我們發現染色體上的強化子（enhancers）⁸對於原病毒基因的表現量是有影響的：當原病毒 DNA 插入染色體的位置越靠近強化子，其基因的表現較那些距離強化子較遠的原病毒相對而言來的高¹⁰。或許你正在納悶什麼是「強化子」¹³？強化子通常是一段長度為 50 到 1500 鹼基對長的 DNA 序列。因為這些 DNA 序列可以和不同的轉錄因子（transcription factors）⁹作用來促進特定基因的表現，因此稱為「強化子」。目前讓我們有些頭痛的問題是：我們仍無法精確地標記出強化子在染色體上的位置。儘管如此，我們發現絕大多數的強化子，其 DNA 序列是帶有特殊的表觀遺傳修飾（epigenetic modifications）¹⁰：最主要的兩種修飾為在組織蛋白 H3 上第 27 號離胺酸（lysine）上所帶有的乙醯化（acetylation）（縮寫為 H3K27ac）以及在組織蛋白 H3 上第 4 號離胺酸上所帶有的單甲基化（mono-methylation）（縮寫為 H3K4me1）。因此，我們可以利用染色質免

疫沉澱-測序 (ChIP-seq) 技術，藉由將染色體上，表現 H3K27ac 和 H3K4me1 訊號較強的位置找出來，來選出強化子可能存在的區域。

在 Chen et al. (2017)¹⁰ 的這篇文章裡，我們首先標記出在染色體上高度表現 H3K27ac 的位置，再選取從此位置的上游及下游各 2500 鹼基對長的區域，用以定義為強化子的區域。我們發現當原病毒落在此區域，其基因表現量較落在此區域外的原病毒基因表現量來得高¹⁰ (圖三)。為了更加確定這個發現，我們利用流式細胞技術 (flow cytometry) 將被病毒感染後的細胞分為兩群：一群是細胞具有綠色螢光蛋白 (green fluorescent protein) 的表現；另一群細胞則不具有綠色螢光蛋白的表現。因為我們的病毒是帶有綠色螢光蛋白基因，因此當細胞被病毒感染後，若帶有綠色螢光蛋白的表現，代表著在這些細胞內的原病毒是具有病毒活性的 (active proviruses)；反之則代表細胞內的原病毒是處於一個休眠的狀態 (latent proviruses)。我們發現到，從不具綠色螢光蛋白表現細胞中所分離出的原病毒，其距離強化子的位置較從具有綠色螢光蛋白表現細胞中所分離出的原病毒來的遠 (圖四)。這個發現呼應了我們之前實驗的結果：當原病毒 DNA 插入染色體的位置越靠近強化子，其病毒基因的表現量則有增強的趨勢。

肆、結語

人類免疫缺陷病毒 DNA 插入宿主染色體的這個步驟在其生命週期中有著關鍵的重要性⁴。人類免疫缺陷病毒唯有將其 DNA 在感染的過程中插入宿主染色體，才可以確保病毒長久存於細胞中並“劫持” (hijack) 宿主細胞的轉錄系統 (transcriptional machinery)，來做出完成其生命週期所必須的種種病毒蛋白質。因此，我們不難想像宿主染色體對於人類免疫缺陷病毒的致病性是會有一定程度的影響。迄今，我們和其他許多的研究都陸續地證明宿主染色體對於人類免疫缺陷病毒基因表現是有影響的¹⁴⁻¹⁷；然而我們也必須說：調控病毒基因表現的機轉是複雜而且多面向的；染色體的因素 (或是說這種類似果蠅位置效應花班的機轉) 只是許多已被發現的機轉中的冰山一角，更遑論那些更多而且尚未被發現的調控機制了。語末，我們提供了一個網站連結，讓有興趣的讀者可以繼續對人類免疫缺陷病毒再做進一步的認識：人類免疫缺陷病毒科學 (THE SCIENCE OF HIV)¹⁸ 計劃的網站。讀者在這個網站中能夠對透過立體模型動畫，對人類免疫缺陷病毒的生命週期以及目前用於治療後天免疫缺乏症候群的策略有更深入的了解。

備註

註 1: <https://www.crg.eu/en>

註 2: <http://www.genomearchitecture.com>

註 3: DamID 技術的全名為 DNA adenine methyltransferase identification (DamID)。因國家教育研究院雙語詞彙、學術名詞暨辭書資訊網尚未對此專有名詞進行收錄，故在此將 DamID 技術暫譯為 DNA 腺嘌呤甲基轉移酶標定技術。此技術是用來尋找蛋白質在染色體上座落的位置。其原理是先將欲研究的蛋白質，利用分子技術，使其帶有 DNA 腺嘌呤甲基轉移酶。當此帶有 DNA 腺嘌呤甲基轉移酶的蛋白質與染色體結合的同時，就會造成鄰近區域 DNA 的腺嘌呤發生甲基化 (adenine methylation) 的現象。因腺嘌呤甲基化在真核細胞 (eukaryote) 中並不是會自然發生的現象，因此利用標定出被甲基化的 DNA 腺嘌呤，即可推測出欲研究的蛋白質與染色體結合的位置。

註 4: 位置效應花斑說明當相同的基因，因存在在染色體的位置不相同，而造成基因表現的不同。目前已經知道造成此現象的主要原因是因為在不同位置的染色體上，帶有不同的表觀遺傳修飾 (epigenetic modification)，進而造成雖然是相同的基因，但其基因的表現卻是不相同。目前已經有許多不同的表觀遺傳修飾陸續被發現，有些表觀遺傳修飾會加強基因的表現，例如在組織蛋白 H3 上第 27 號離胺酸 (lysine) 上所帶有的乙醯化

(acetylation) (縮寫為 H3K27ac); 有些則會抑制基因的表現，例如在組織蛋白 H3 上第 9 號離胺酸上所帶有的三甲基化 (tri-methylation) (縮寫為 H3K9me3)。

註 5: 按國家教育研究院雙語詞彙、學術名詞暨辭書資訊網所收錄的學術名詞 (學術名詞—生命科學名詞)，在此將 latency 翻譯為潛伏態。潛伏態即指在病毒感染的過程中，存有一特定的時期，其病毒的基因表現量處於在目前技術可檢測出的病毒量的標準值之下，如同處於一個休眠的狀態中，故稱“潛伏”。

註 6: 因國家教育研究院雙語詞彙、學術名詞暨辭書資訊網尚未對此專有名詞進行收錄，故在此我們並未對此技術名稱進行翻譯。此技術的優勢在於能夠同時將數以千計的報導基因 (reporter gene) 送入細胞內來進行實驗，並利用生物資訊分析 (bioinformatic analysis) 技術來加以區別，而獲得大數據 (Big Data)。因有別於傳統上，單次送入單一報導基因的實驗，這樣的技術因此被稱為高通量 (high-throughput) 技術。

註 7: 按國家教育研究院雙語詞彙、學術名詞暨辭書資訊網所收錄的學術名詞 (醫學名詞—醫事檢驗名詞)，在此將 high-throughput sequencing 翻譯為高通量定序。此技術的優勢是在於能

夠在短時間內，快速而有效地處理大量的樣品。

註 8: 按國家教育研究院雙語詞彙、學術名詞暨辭書資訊網所收錄的學術名詞（學術名詞—生命科學名詞），在此將 **enhancer** 翻譯為強化子。

註 9: 按國家教育研究院雙語詞彙、學術名詞暨辭書資訊網所收錄的學術名詞（學術名詞—生命科學名詞），在此將 **transcription factor** 翻譯為轉錄因子。轉錄因子是指能夠結合在某基因上游特異核苷酸序列上的蛋白質，來調控其基因的表現。

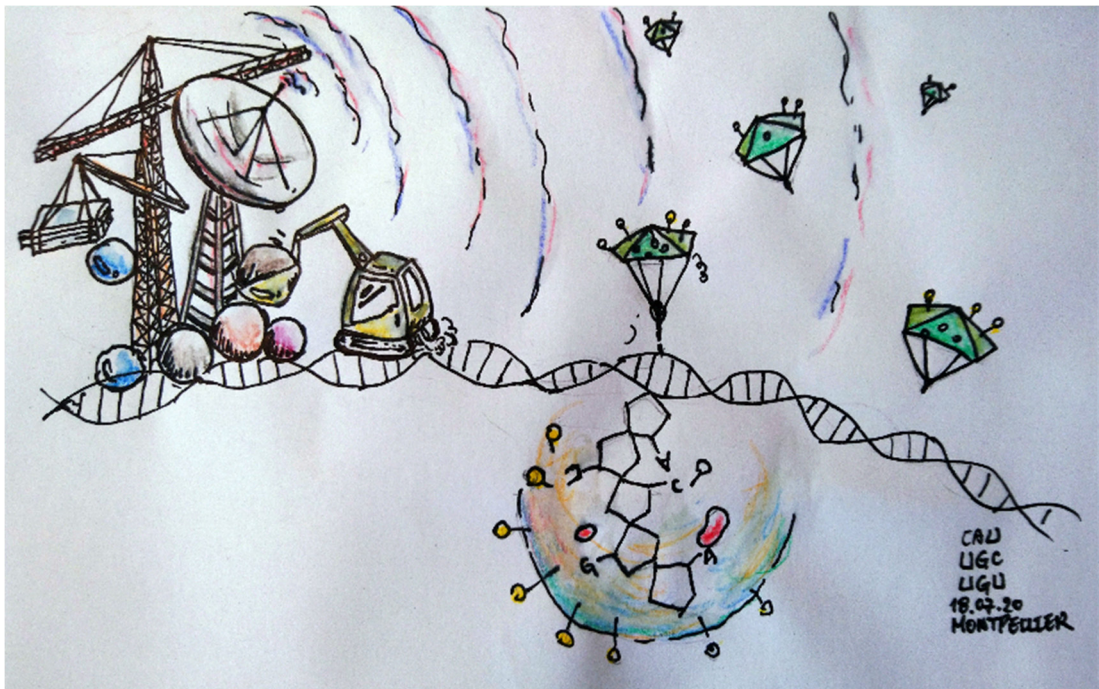
註 10: 因國家教育研究院雙語詞彙、學術名詞暨辭書資訊網尚未對此專有名詞，在醫學、生物相關領域內，進行收錄。故在此借用國家教育研究院雙語詞彙、學術名詞暨辭書資訊網所收錄的學術名詞（學術名詞—電機工程），將 **epigenetic modification** 翻譯為表觀遺傳修飾。

參考文獻

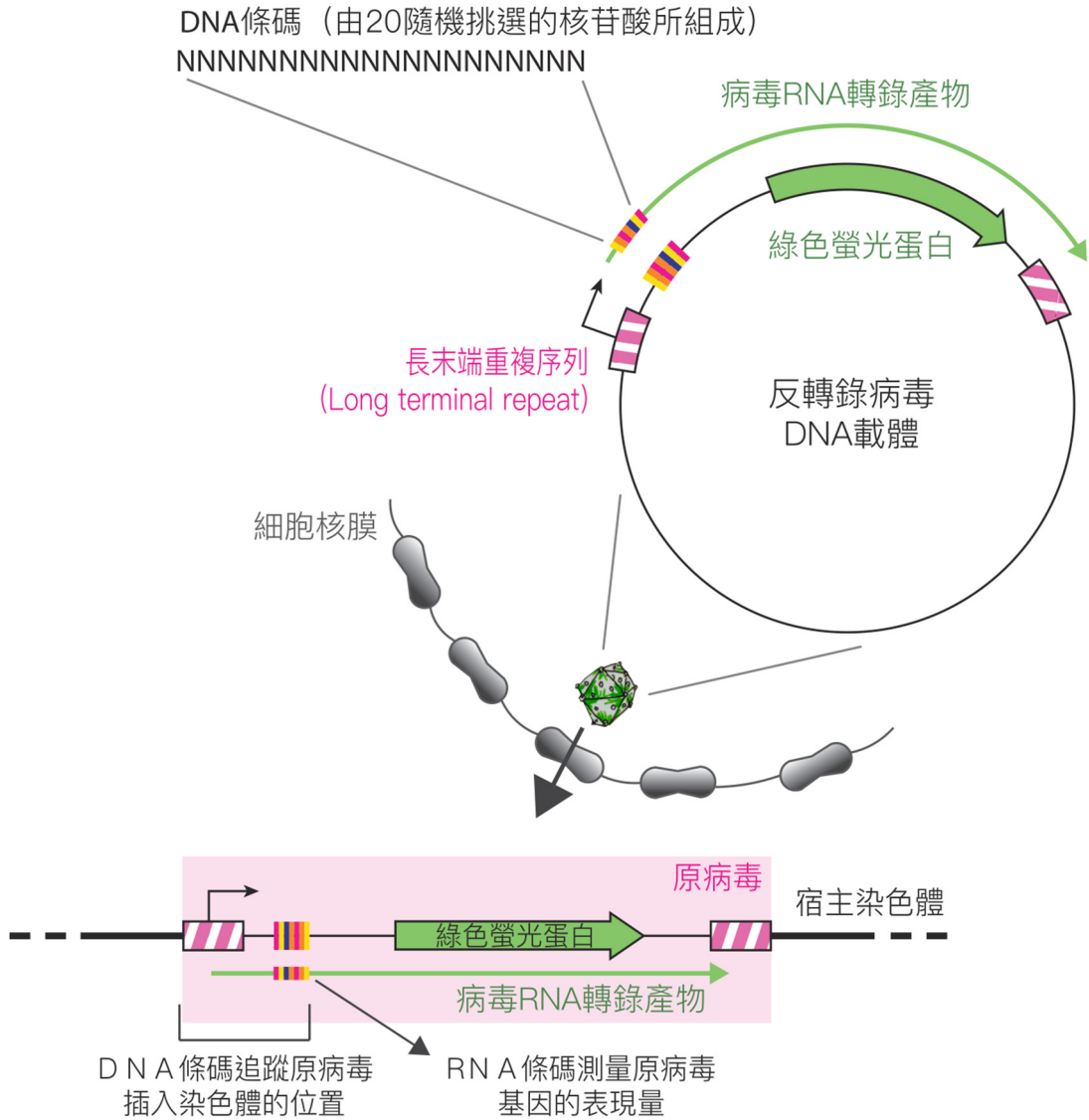
1. van Steensel, B., Delrow, J. & Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* **27**, 304–308.
2. Filion, G. J. *et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224.
3. Muller, H. J. (1930). Types of visible variations induced by X-rays in *Drosophila*. *Journal of Genetics* vol. 22 299–334.
4. Pierson, T., McArthur, J. & Siliciano, R. F. (2000). Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annu. Rev. Immunol.* **18**, 665–708.
5. Schröder, A. R. W. *et al.* (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529.
6. Lewinski, M. K. *et al.* (2005). Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**, 6610–6619.
7. Marini, B. *et al.* (2015). Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521**, 227–231.
8. Jordan, A., Defechereux, P. & Verdin, E. (2001). The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* **20**, 1726–1738.
9. Jordan, A., Bisgrove, D. & Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**, 1868–1877.
10. Chen, H.-C., Martinez, J. P., Zorita, E., Meyerhans, A. & Filion, G. J. (2017). Position effects influence HIV latency reversal. *Nat. Struct. Mol. Biol.* **24**, 47–54.
11. Chen, H.-C., Zorita, E. & Filion, G. J. (2018). Using Barcoded HIV Ensembles (B-HIVE) for Single Provirus Transcriptomics. *Curr. Protoc. Mol. Biol.* **122**, e56.
12. Akhtar, W. *et al.* (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**, 914–927.
13. Blackwood, E. M. & Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60–63.
14. Battivelli, E. *et al.* (2018). Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4 T cells. *Elife* **7**.
15. Vansant, G. *et al.* (2019).

- Impact of LEDGIN treatment during virus production on residual HIV-1 transcription. *Retrovirology* **16**, 8.
16. Lindqvist, B., Svensson Akusjärvi, S., Sönnnerborg, A., Dimitriou, M. & Svensson, J. P. (2020). Chromatin maturation of the HIV-1 provirus in primary resting CD4+ T cells. *PLoS Pathog.* **16**, e1008264.
17. Vansant, G. *et al.* (2020). The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res.* doi:10.1093/nar/gkaa536.
18. The Science of HIV project. The Animation Lab at the University of Utah. From <https://scienceofhiv.org/wp/>.

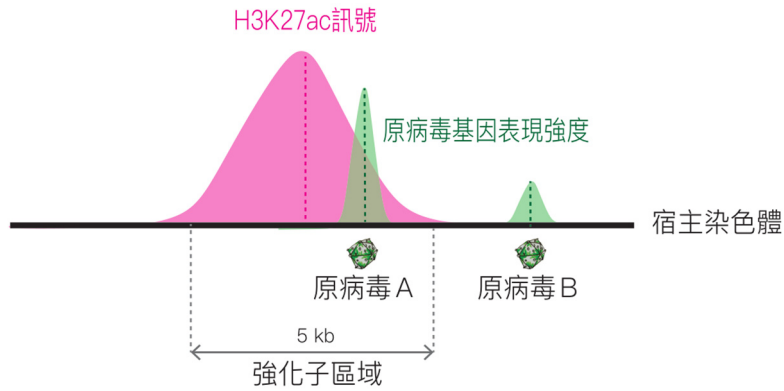
圖說



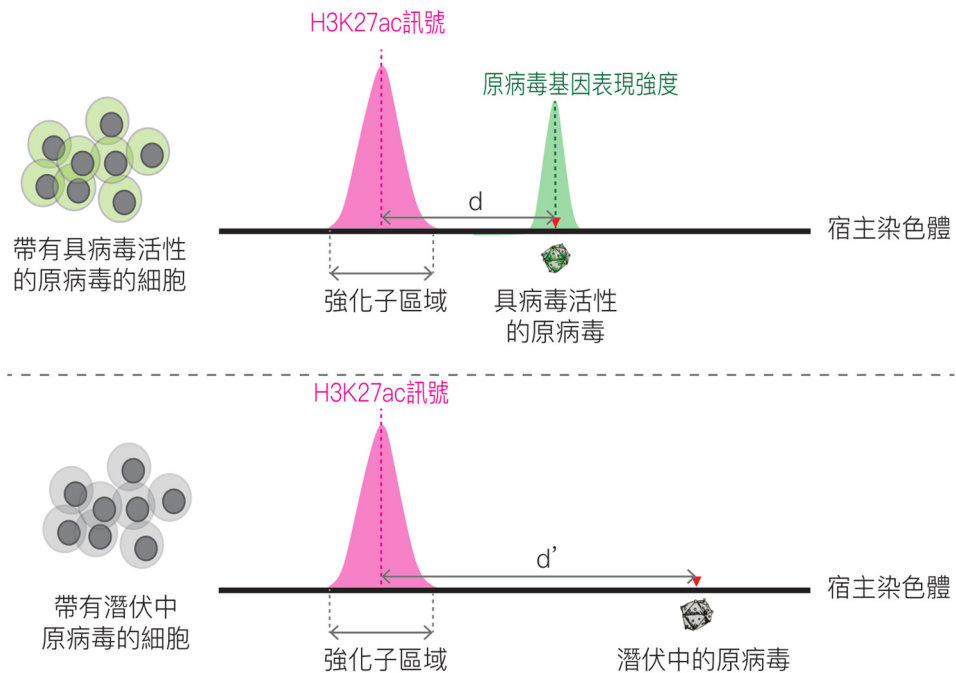
圖一：人類免疫缺陷病毒 DNA 插入染色體的位置影響病毒在宿主細胞內的命運。原病毒在染色體上的位置不同，是可以決定病毒將進入休眠的狀態或是繼續表達病毒蛋白質，來完成其生命週期。在此卡通圖中，我們用傘兵代表感染細胞後的人類免疫缺陷病毒；雷達基地台則代表染色體上強化子所在的區域。



圖二：B-HIVE 技術的原理。我們將一段帶有二十個隨機挑選的核苷酸所組成的 DNA 序列 (DNA 條碼) 放入人類免疫缺陷病毒 5' 端長末端重複序列 (5' long terminal repeat) 後方，藉由追蹤 DNA 條碼所在染色體上的位置來標定出原病毒 DNA 插入染色體的位置。同時，利用測量條碼 RNA 的表現量，我們則可知道相對應原病毒基因的表現量。



圖三：染色體上的強化子影響原病毒基因的表現。當原病毒 DNA 插入染色體的位置越靠近強化子，其病毒基因表現則有越強的趨勢。桃紅色區域表示經由 H3K27ac 訊號標記出的強化子區域。綠色區域用以示意原病毒基因表現的強度。



圖四：潛伏中原病毒 DNA 插入染色體的位置距離強化子較遠。當比較具有病毒活性的原病毒和潛伏中的原病毒，其 DNA 插入染色體的位置和染色體上的強化子的距離，我們發現到潛伏中的原病毒（距離以 d' 表示）距離染色體上強化子的區域，較具病毒活性的原病毒（距離以 d 表示）來得遠。紅色倒三角形代表原病毒 DNA 插入染色體的位置。