[Hierarchical Ridge Regression for Incorporating Prior Information in Genomic Studies](https://jds-online.org/journal/JDS/article/1262/info)

基因組數據通常是高維的，高維回歸方法需要正則化(Ridge Regression)，也就是添加迴歸係數的懲罰項。然而，增加此項會需要在模型複雜性（偏差）和模型穩定性（方差）之間的取捨。前人想到可以利用meta feature來調整，讓上述問題簡化。

之所以要做出修改，是因為比一般機器學習應用場景feature多更多，直接使用係數作為loss function會增加太多複雜度，因此使用meta/ subjective 交互使用的策略降低複雜度，同時減少穩定性的trade off

這邊提出兩層的normalization，第一層一樣在subject-level features上面做回歸。第二層模擬meta feature對subject-level features平均值的影響。如此就可以用一般迴歸分析，但又不易產生過於複雜的模型。

[Interpretable machine learning prediction of all-cause mortality](https://pubmed.ncbi.nlm.nih.gov/36204043/)

eXplainable AI = XAI

impact

在一般主題裡，機器學習的結果就是黑盒。我們無從得知判定的原因。但再生醫領域中，許多判斷會讓病人需要付出很大的成本，甚至是生命，因此希望可以再做決定的時候給病人原因，讓治療過程更為令人安心。這邊使用gradient boosted tree，產生tree explainer來解釋。

文中，他們取得對未來死亡率提供高度信息的風險預測因子，利用靈活的模型捕獲非線性關係(例如: 風險預測因子的“拐點”提供的重要資料)。再利用xai重新回推，理解那些feature 是對死亡重要的。

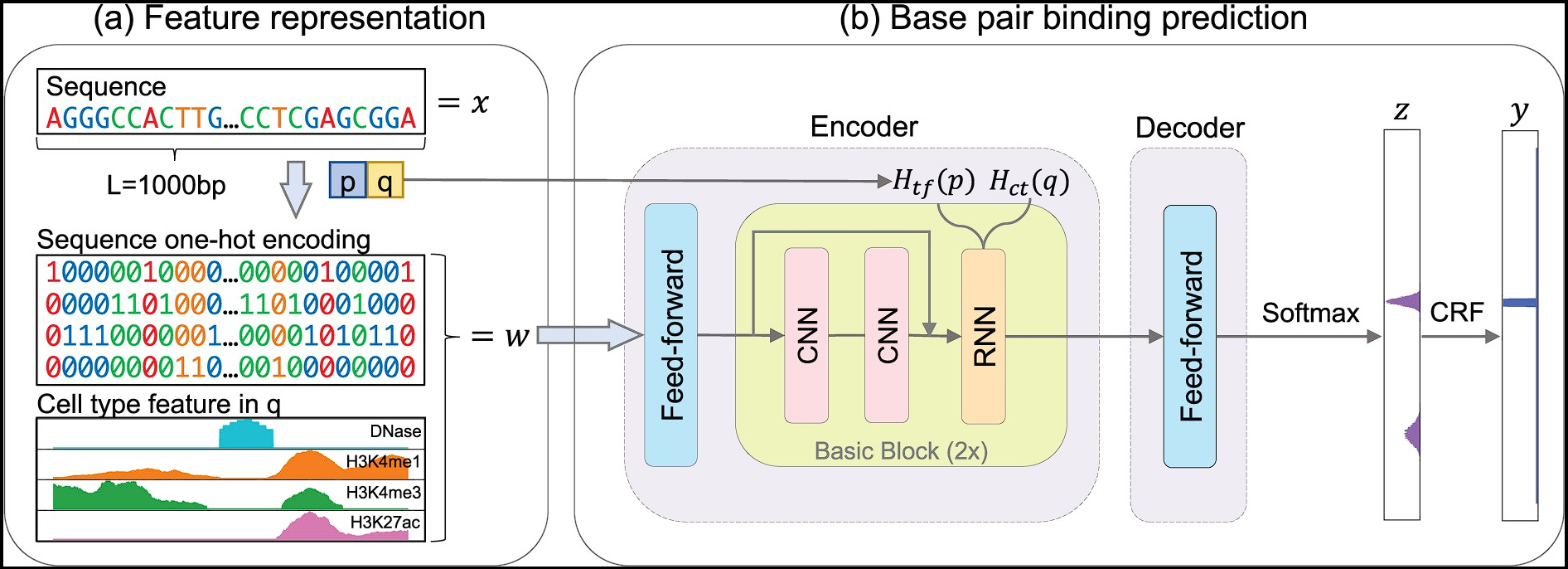
此篇文章結果比死亡風險評分和生理年齡都更準確，有一定研究價值。

有新的 對死亡重要的原因。

[**NetTIME: a multitask and base-pair resolution framework for improved transcription factor binding site prediction**](https://academic.oup.com/bioinformatics/article/38/20/4762/6673901)

TF binding prediction是指給定特定dna，蛋白質是否會彼此結合，某一特定細胞種類。此題目一直是生醫領域的熱門主題。但一直以都只能用特定cell type來預測 (因為不同細胞種類情形相差很大)這個模型透過分析cell type feature 來擬合不同type的情況。利用one-hot encoding 分析，給auto encoder 做CNN。

重點可以多放在如何分析cell type feature (利用幾個chip -seq做feature)這件事情合理性、好不好等等。



不要的:

[A machine learning approach utilizing DNA methylation as an accurate classifier of COVID-19 disease severity](https://pubmed.ncbi.nlm.nih.gov/36261477/)

使用jadbio 自動化生醫機器學習。它包含了（嶺邏輯回歸）、非標準線性模型（線性 SVM）和非線性模型（決策樹、隨機森林、具有非線性內核的 SVM）等等的方法。文章中特別強調，在資訊過多的甲基化系統，使用feature selection algorithm 可以有效減少不必要的資料。

(需要替換原因：使用自動機器學習平台，有點太過簡單)

<https://reurl.cc/NGR936>