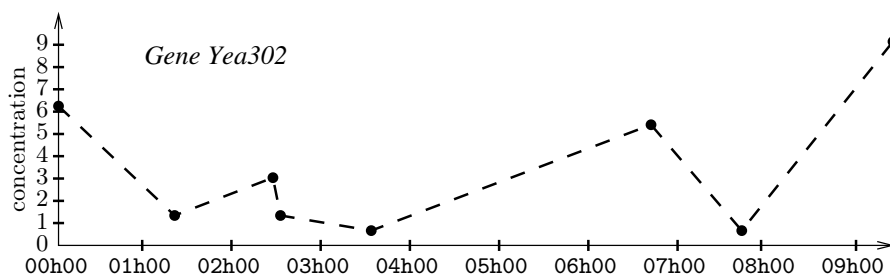


Automates, dictionnaires et listes

Le but de ce TD est de repérer les gènes covariants dans une liste de profils d'expression similaire à celles manipulées dans les TD précédents. Pour cela on veut utiliser la fonction `coexprime` du TD n°4.

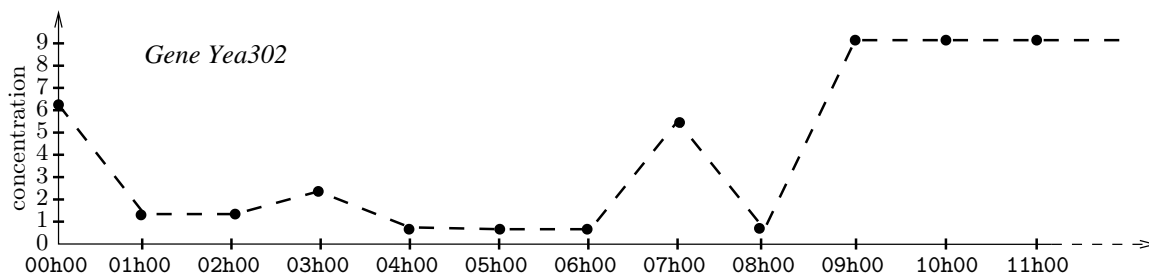
La difficulté vient du fait que nos profils d'expressions ne sont pas tous mesurés aux mêmes horaires d'un gène à l'autre, alors que la fonction `coexprime` prend en entrée deux listes de niveaux d'expressions qui représentent des mesures faites aux *mêmes heures*.

Il faut donc commencer par « normaliser » chaque profil d'expression pour avoir un niveau d'expression « théorique » pris aux mêmes heures pour chaque gène. On va calculer ce niveau théorique pour chaque heure exacte. Par exemple, le dictionnaire `yea` du TD précédent représente le profil d'expression ci-dessous :



```
{ "nom": "Yea302", "00h00":6.3, "01h21":1.4, "02h32":3.0, "02h40":1.4, "03h35":0.7,
"06h45":5.6, "07h50":0.7, "09h12":9.0 }
```

On voit bien que les mesures ne sont pas prises à chaque heure. On va donc « lisser » ce profil d'expression comme suit :



Ce profil d'expression « théorique » est obtenu selon le principe de l'exercice ci-dessous.

Exercice 1 : Écrivez une fonction `normalise` qui prend en entrée un dictionnaire `d` représentant un profil d'expression quelconque et le « lisse » en donnant une liste de mesures « théoriques » avec une mesure pour chaque heure ("00h00", "01h00", ..., "23h00"). On procède de la manière suivante :

- On fournit un message d'erreur si le dictionnaire de départ ne possède aucune mesure comprise entre les temps "00h00" et "00h30", sinon la première mesure de la liste (temps 0) est la moyenne des mesures comprises entre ces deux temps.
- Ensuite, pour chaque heure `X`, on associe la moyenne des valeurs du dictionnaire de départ entre les temps "`(X-1)h30`" et "`Xh30`".

Il peut arriver cependant qu'il n'y ait aucune mesure entre ces deux temps. Dans ce cas, on reporte la moyenne précédente (celle de `X-1`).

- La liste retournée par la fonction `normalise` contiendra 24 mesures théoriques (de "00h00" à "23h00").

INDICATION : *Il faut raisonner en utilisant la notion d'automate...* et on utilisera les fonctions `horloge` et `horaire` du TD précédent.

Note : cet algorithme de lissage est peu performant, comme on le voit en comparant les courbes données en exemple ; n'hésitez pas à en inventer de nouvelles versions plus crédibles s'il vous reste du temps ou en exercice à la maison.

Exercice 2 : Écrivez la fonction `cluster` qui prend en entrées :

- d'une part, un dictionnaire `ref` pour un gène de référence,
- et d'autre part, une liste de dictionnaires `ld` représentant une liste de profils d'expression, et fournit en sortie la liste des noms des gènes co-exprimés avec `ref`.