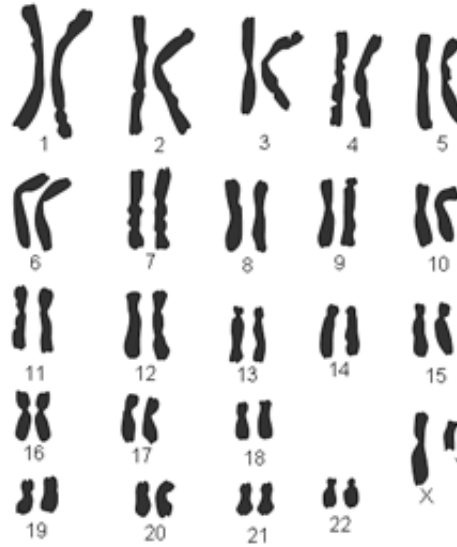




Bioinformatique et génomique des cancers

alain.viari@inria.fr



caryotype humain
normal (homme)

23 paires de chr.

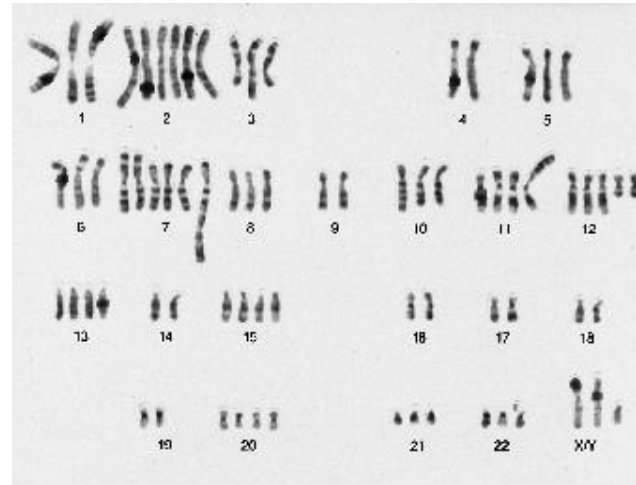
génomome = $3 \cdot 10^9$ pdb
(x2)

Theodor Boveri (1862-1915)

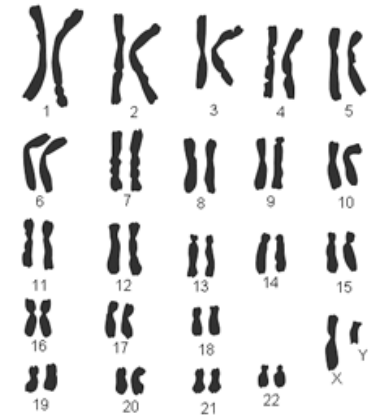
- chromosomes = support matériel de l'hérédité
(+ Walter Sutton)
- hypothèse: cellule tumorale -> dérèglement
des chromosomes (1902)



tumeur



tissu normal



Theodor Boveri (1862-1915)

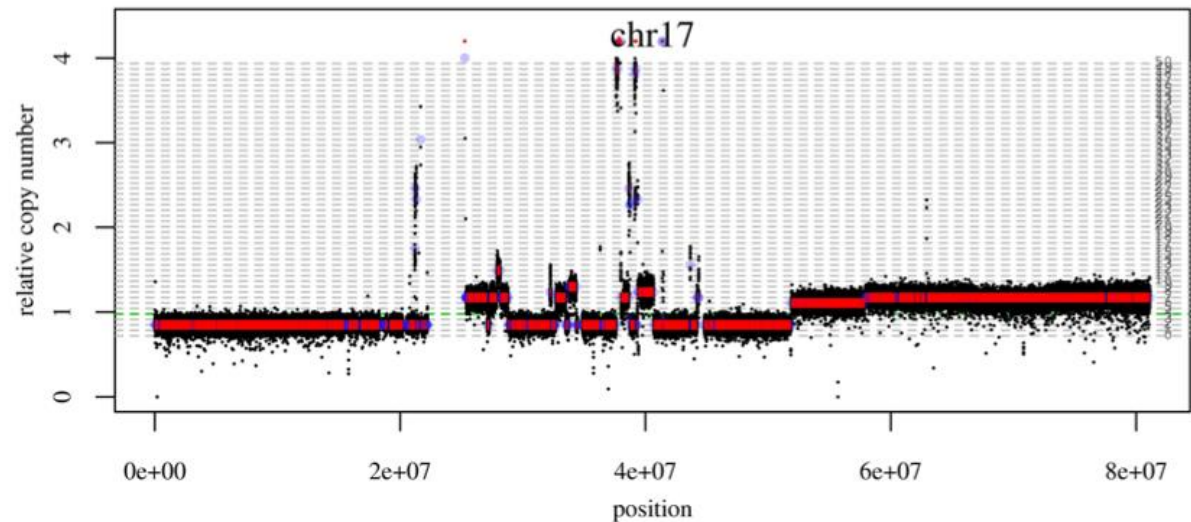
- chromosomes = support matériel de l'hérédité (+ Walter Sutton)
- hypothèse: cellule tumorale <- dérèglement des chromosomes (1902)



Theodor Boveri (1862-1915)

- chromosomes = support matériel de l'hérédité (+ Walter Sutton)
- hypothèse: cellule tumorale <- dérèglement des chromosomes (1902)

ErbB2

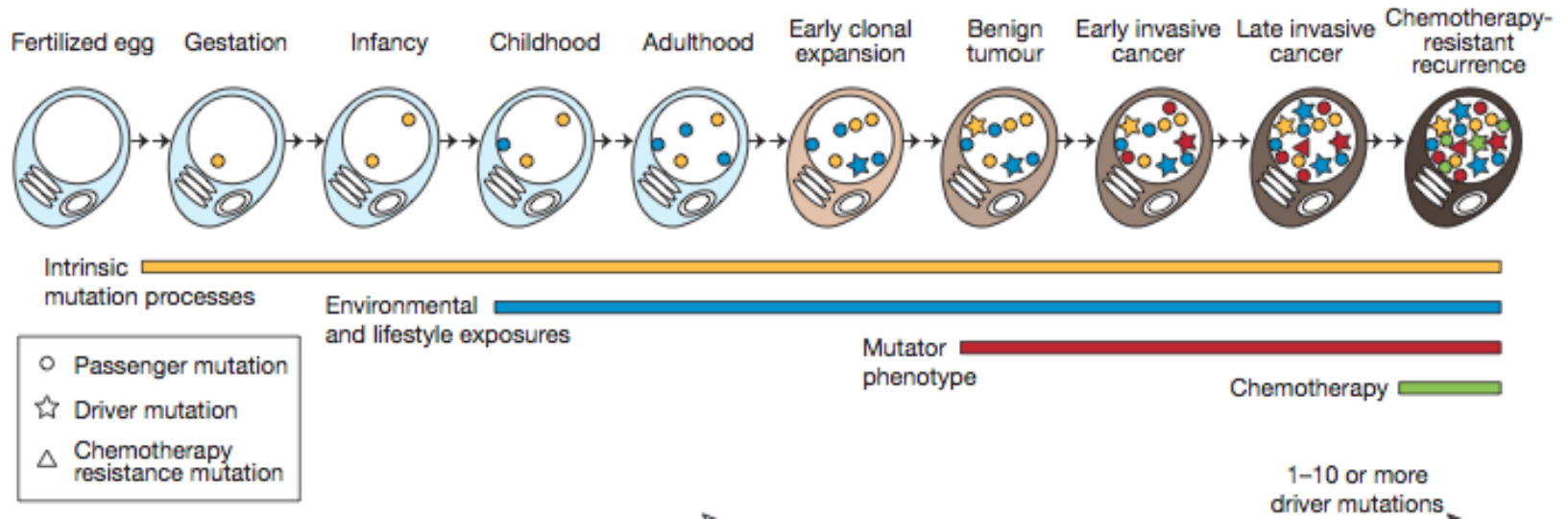


ex: cancer du sein (HER2+)
chr 17

cancer: altérations somatiques

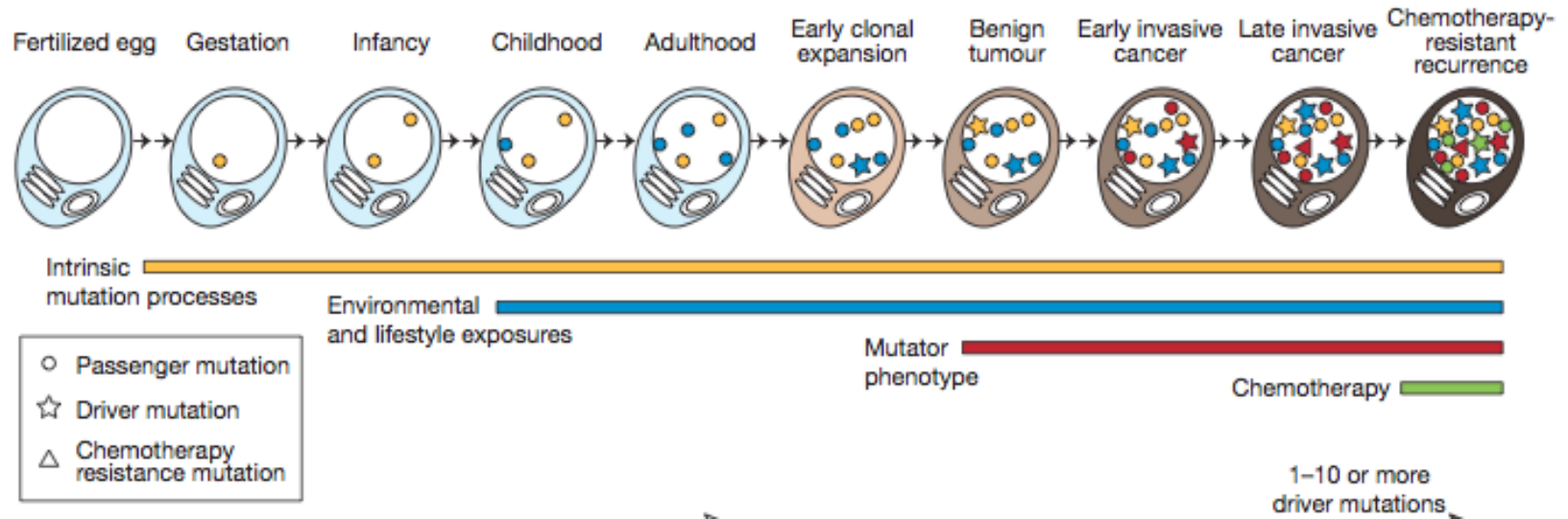
origine du dérèglement : altérations **somatiques** (\neq germinale)

from Stratton et al. 2009



origine du dérèglement : altérations **somatiques** (\neq germinales (constitutionnelles))

from Stratton et al. 2009



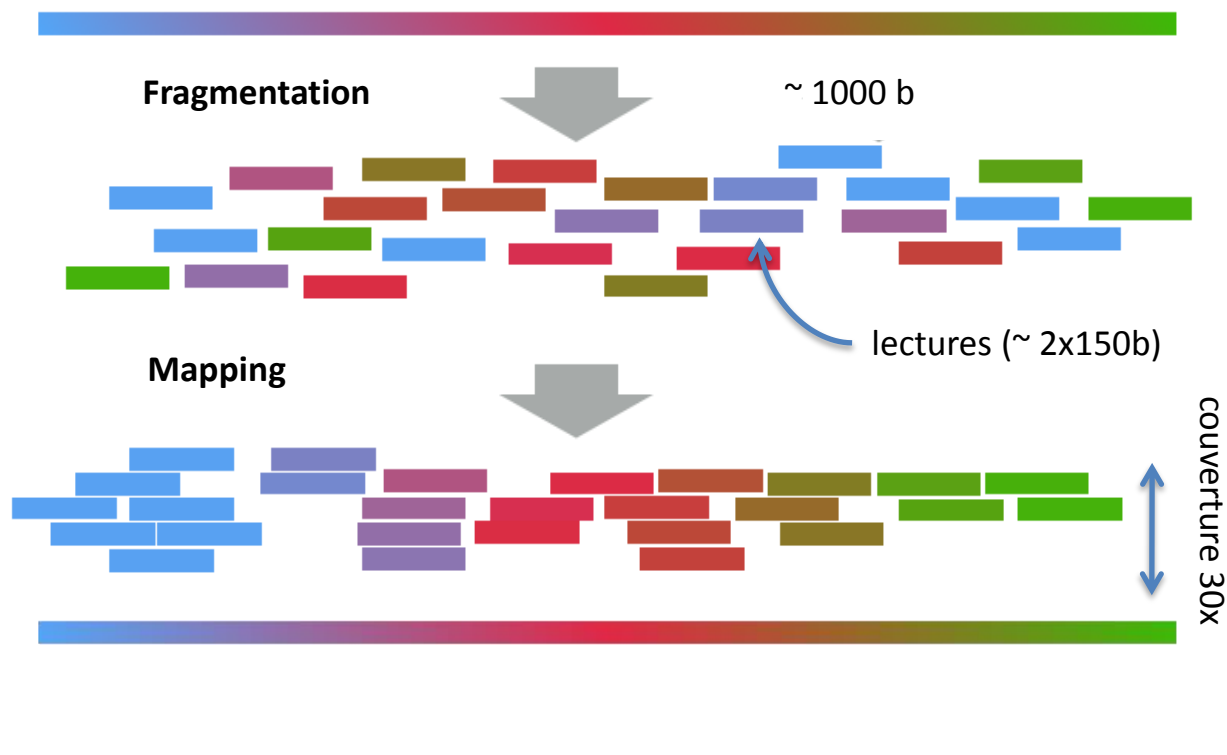
driver : à l'origine du dérèglement

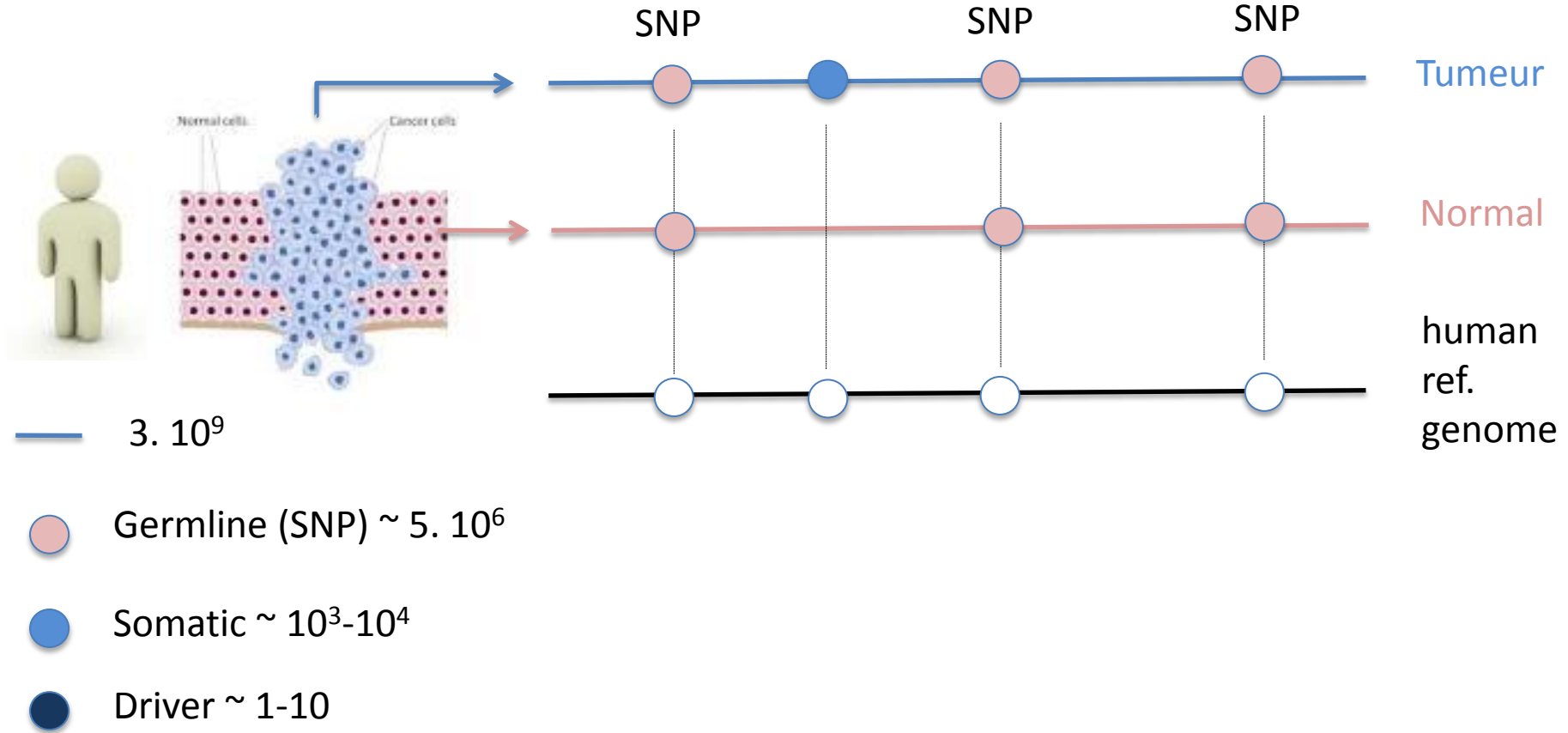
passenger: conséquence du dérèglement

(+ background)

- oncogènes amplifiés/activés (ex: ERBB2)
- gènes suppresseurs de tumeurs délétés/mutés/réprimés (ex: P53)

NGS
Next Generation
Sequencing





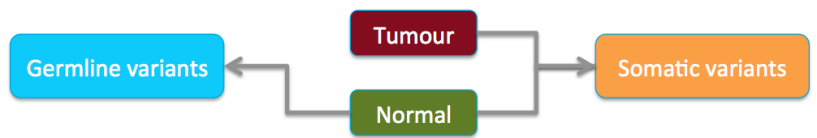
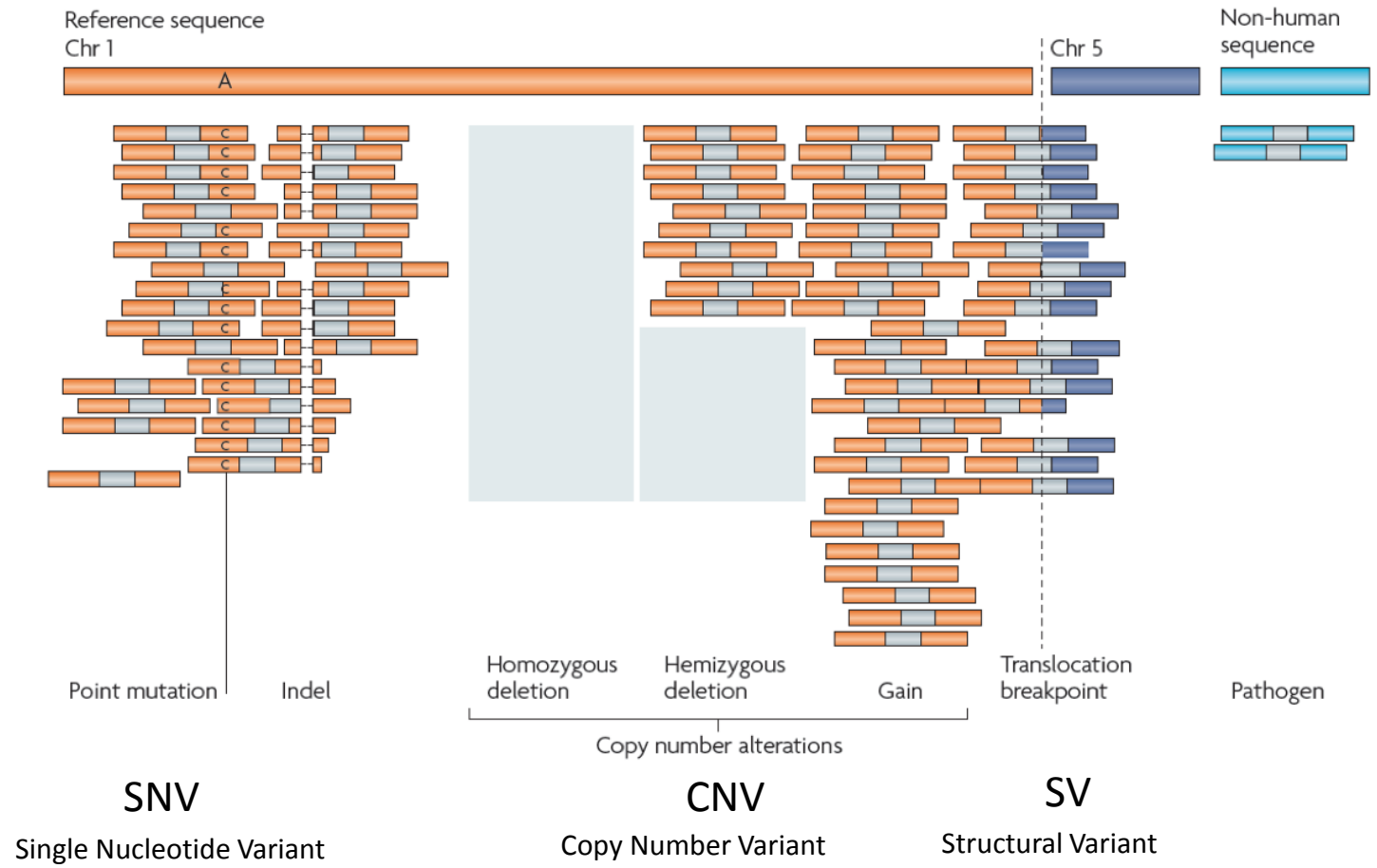
note: il existe aussi des mutations germinales qui confèrent une susceptibilité (ex BRCA1/2)

=> **mutations constitutionnelles**

- vitesse : 5 à 10 WGS (30x) / jour (1 NovaSeq)
- coût : ~ 1500 euros / génome (30x)

NGS en cancérologie : quel types d'informations ?

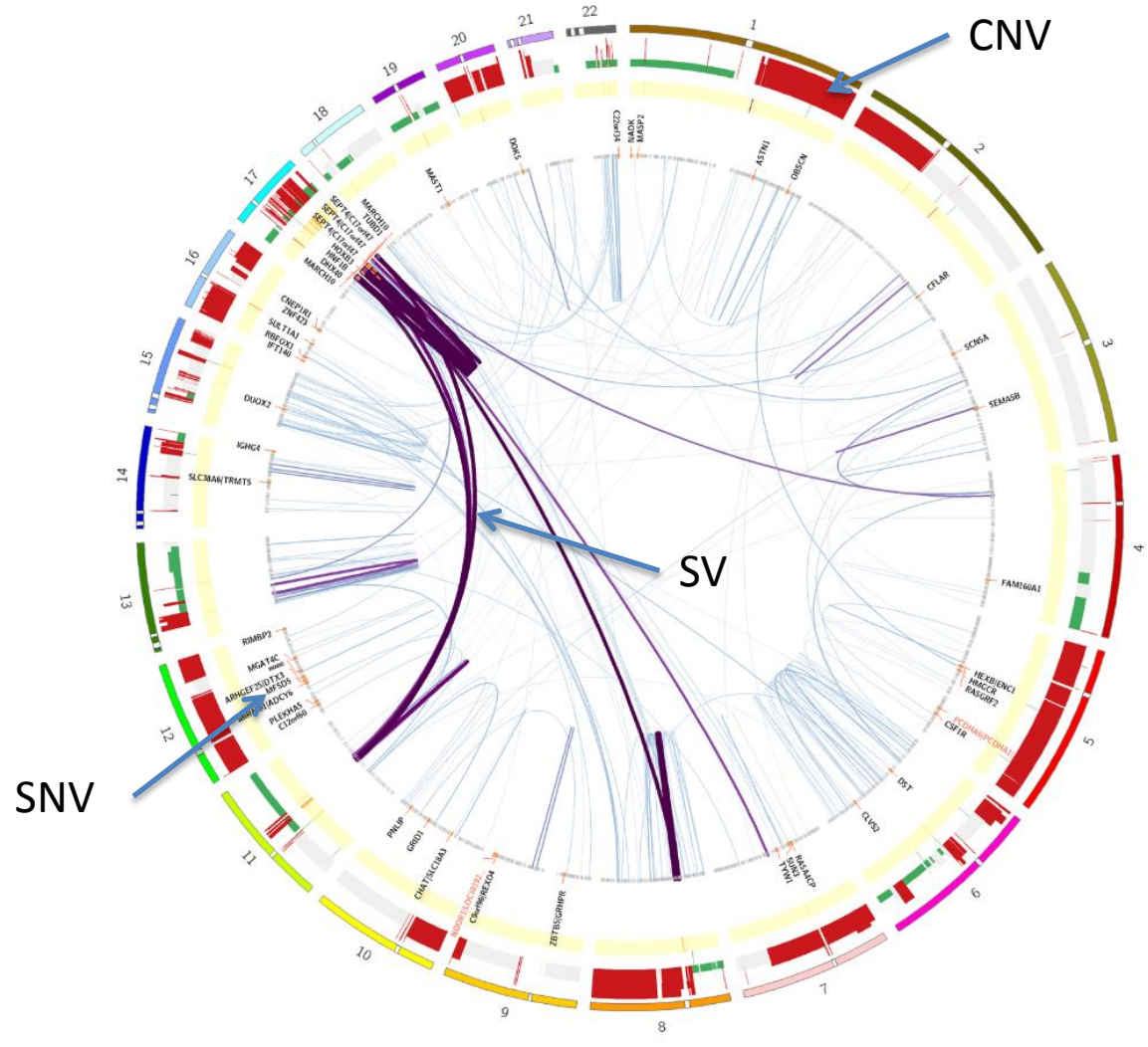
from Meyerson et al. 2010



NGS en cancérologie : quel types d'informations ?

“carte d'identité génomique”
par patient

- comprendre les relations entre les altérations
- comprendre les processus à l'oeuvre
- prédire de nouveaux biomarqueurs



cancer du sein HER2+

Cadre Recherche versus Diagnostic/Thérapeutique

Recherche



cohorte

- altérations récurrentes
- recherche de biomarqueurs
- compréhension des mécanismes

Diagnostic



patient
+médecin

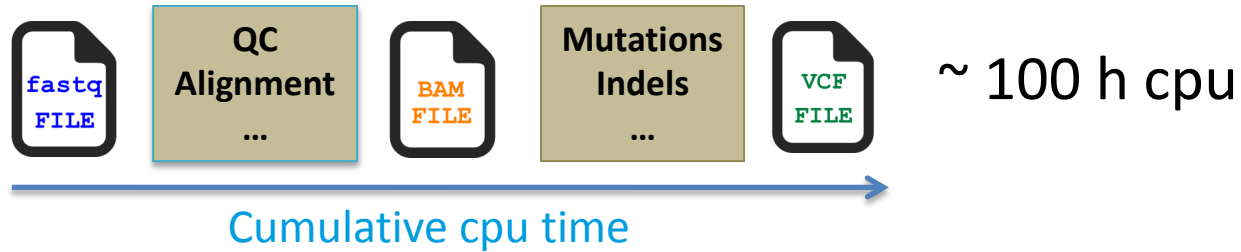
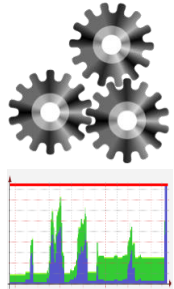
- altérations connues, validées et **actionnables**
- prise en charge thérapeutique

Next Generation Sequencing

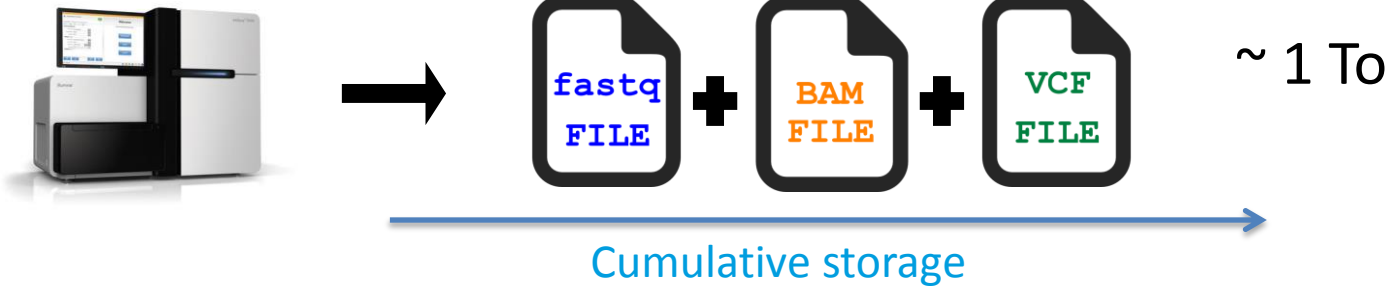
| ADN | | | ARN |
|---|--|------------------------------------|---|
| Genome complet (WGS) | Exome (WES) | Ciblé | RNASeq |
| (2x) 3 Gb x 50 | 3% génome x 1000 | quelques gènes x 1000 (ou plus) | transcrits expr. |
| mutations somatiques variants structuraux signatures mutationnelles | mutations somatiques et constitutionnelles | | variants épissage expression des gènes |
| Recherche | | Diagnostic/Soin | Recherche |



par patient (normal+tumoral en ~ 50x)



Hiseq 2000



=> nécessité d'une **architecture dédiée** (cluster de calcul, cloud)

ex: OICR (2016) ~ 6000 cœurs / 2.5 Po stockage



Cadre soin : des contraintes supplémentaires

- altérations connues, validées et actionnables

- multiplicité des acteurs

(clinicien, anatomo-pathologiste, biologiste, bioinformaticien)

=> nécessité d'une orchestration du processus

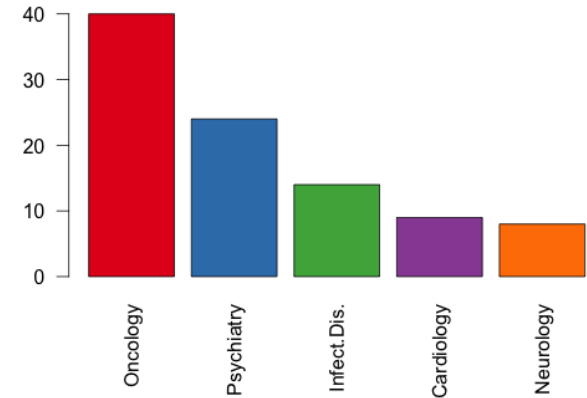
- lisibilité des résultats par le prescripteur

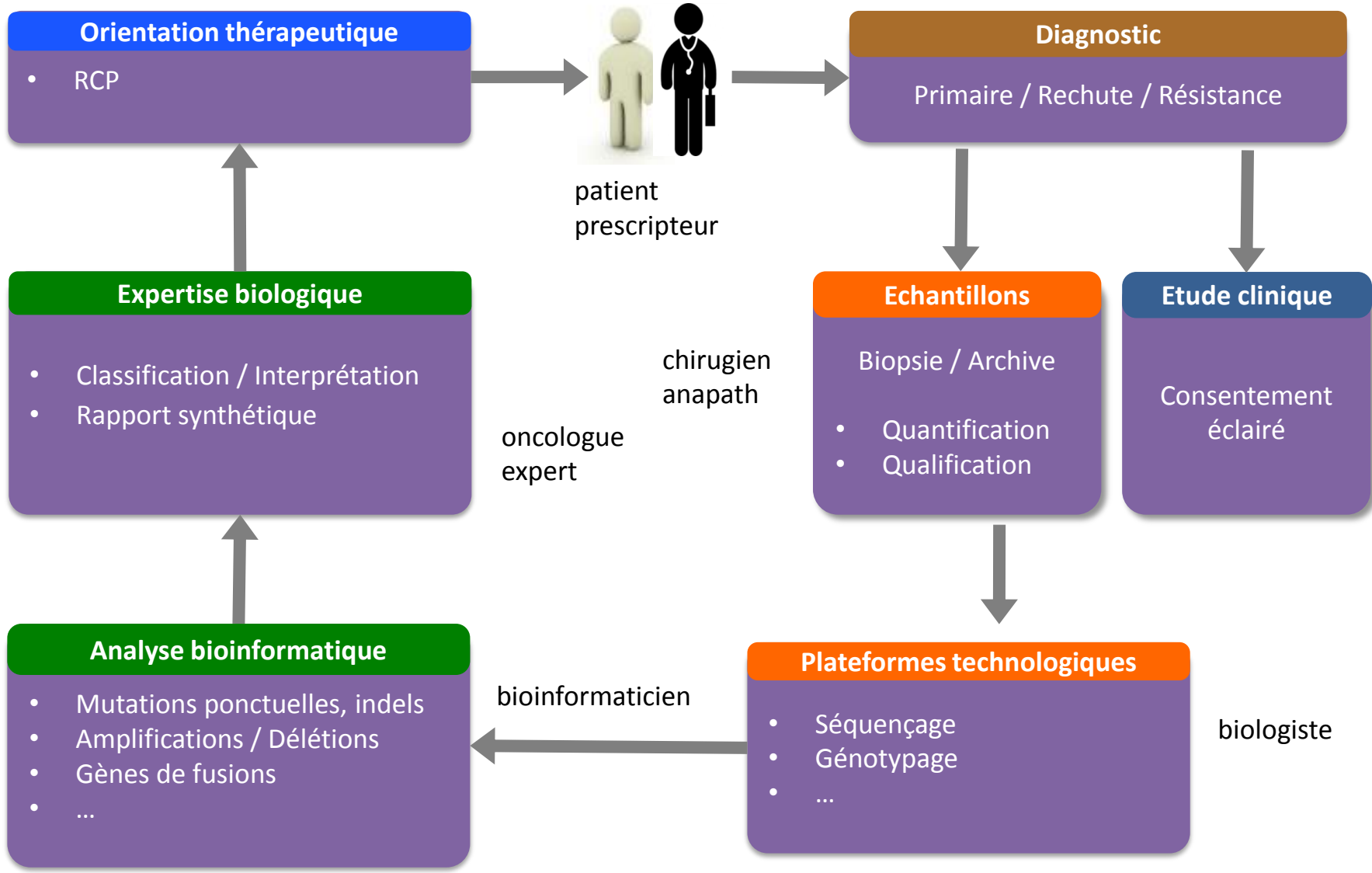
- pas de droit au « bug »

=> robustesse du software, accréditation et sécurité des données

- contraintes réglementaires et légales

FDA Pharmacogenomic Biomarkers
in Drug Labeling





=> le système d'information est central **et critique** pour l'orchestration

| Logiciel | |
|-----------------------------|----------------------|
| Tests | Preuve/Certification |
| Unitaires Fonctionnels | Preuve de programme |



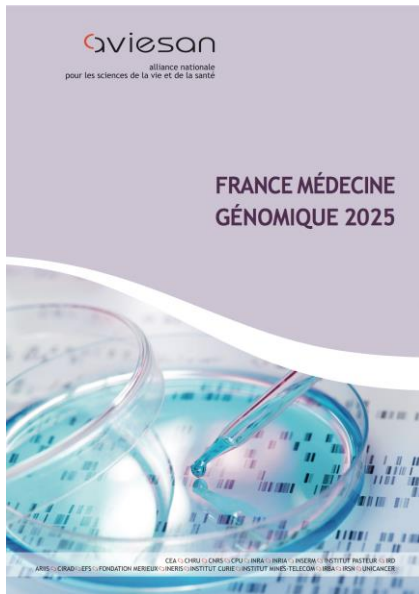
=> les conséquences d'un bug logiciel sont **très** graves

=> les pipelines de traitement utilisés actuellement (même en diagnostic) sont :

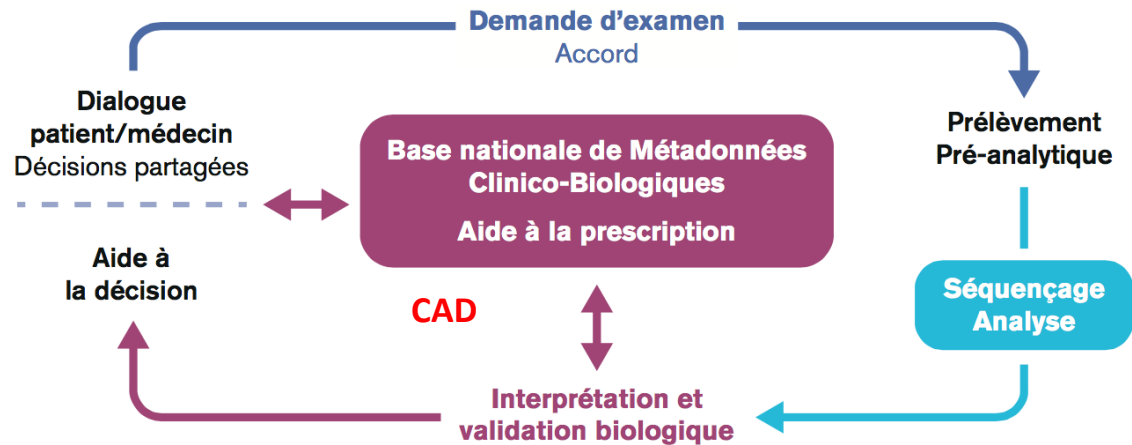
- non testés
- pour certains difficilement testables
- non isolés du système d'exploitation



=> difficulté d'identification des responsabilités en cas de problème



Parcours générique : demande d'examen, préanalytique, analytique, interprétation biologique, rendu



Plan FMG-2025 – Soins – WGS/WES/RNASeq – 200.000 génomes/an (cancer, MR, MC)

- les contraintes du soin + la volumétrie de la recherche
- + un formidable défi pour les cliniciens/biologistes/bioinformaticiens

la collection centralisée (CAD) des métadonnées clinico-biologiques et des données de séquence va être déterminante...



Merci de votre attention

remerciements à la plateforme bioinformatique
« Gilles Thomas » Lyon – CLB - INCa

the « Gilles Thomas » bioinformatics facility in Lyon

