GÉNOMIQUE DU CANCER
Origine du cancer

cancer = maladie du génome
accumulation de mutations dans les cellules souches
⇒ cellules somatiques aberrantes (évitant la mort cellulaire)

Figure 1. The structure of epithelial tissue compartments influences the accumulation of somatic mutations and progression to cancer. This figure illustrates a crypt, the compartmental unit of colon tissue. The stem cells reside at the base of the crypt. Each stem cell division typically gives rise to one stem cell that remains at the crypt base and one transit cell that moves up. The transit cell then divides several times, pushing the cells above toward the colon surface, where the surface cells undergo apoptosis and are shed. The stem cells form the only long-lived cell lineages, from which other crypt cells derive. Thus, cancer progression mostly follows the accumulation of mutations to stem cell lineages.

Figure 2. The phases of cellular growth in epithelial tissues. Cell populations expand exponentially during development, shown by a branching phase of division. At the end of development, stem cells differentiate in each tissue compartment. Stem cells renew each compartment by dividing to form a nearly linear cellular history—each stem cell division gives rise to one daughter stem cell that continues to renew the tissue and one daughter transit cell that divides rapidly to produce a short-lived transit lineage that fills the tissue.

Mutations somatiques

genre de mutations dans le génome d’une cellule cancéreuse :
★ mutations de point (missense, nonsense) et petits indels (frame shift)
★ réarrangements (fusions, translocations, duplications), changement de ploïdie
★ épigénomique (hyperméthylation)

rôle des mutations :
★ **driver** : avantage pour la lignée clonale (sélection positive)
★ **passenger** : «auto-stoppeur» sans avantage particulier

Analyse par séquençage

On peut détecter les modifications du génome par séquençage :

to be involved in tumorigenesis, were revealed for a cancer. The fact that the most frequently mutated genes they observed, \(\text{APC}\), \(\text{TP53}\), and \(\text{KRAS}\) for colon cancer and \(\text{TP53}\) for breast cancer, recapitulated what was already known, validated the approach and paved the way for expanded application of genome-scale sequencing.

The introduction of DNA sequence enrichment technologies from NimbleGen and Agilent (Albert et al. 2007; Gnirke et al. 2009) enabled WES on large scales. WES has additional advantages over WGS in that the average depth of coverage is about fivefold greater, and the cost of sequencing, data processing and storage are all much less. Given the relative tractability of interpreting variation in the coding sequence compared to intergenic or intronic mutations, the period between 2004 and 2013 has seen a profusion of tumor types analyzed in large cohorts (100–500 patients), mainly by WES (see http://www.sanger.ac.uk/genetics/CGP/cosmic/papers/ for a comprehensive listing). WGS for a variety of tumors has also been reported and, in spite of the smaller numbers of patients, has led to surprising insights into cancer biology, based largely on analysis of structural variation in tumor genomes. Using WGS, genetic alterations observed in the DNA of the cancer cell span six orders of magnitude, from single-base point mutations to chromosome-scale amplification, using different modes of sequence analysis (see Chin et al. 2011) available today.

With these tools in hand, The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/), the Cancer Genome Project (http://www.sanger.ac.uk/genetics/CGP/), the International Cancer Genome Consortium (ICGC) (Hudson et al. 2010), Therapeutically Applicable Research to Generate Effective Treatments (http://target.cancer.gov/), and other privately funded large-scale projects (Downing et al. 2012) began in earnest to systematically catalog all the mutations in a wide variety of adult and pediatric cancers (see Garraway and Lander 2013 for a recent tally of large-scale projects).

WGS and WES sequencing have been augmented by cDNA sequencing (referred to as RNA-seq) to explore alterations to the transcriptome. RNA-seq provides not only gene expression levels, but also aberrant splicing, chimeric gene fusion transcripts characteristic of cancer cells and expressed somatic mutations (Bainbridge et al. 2006; Dong et al. 2009; Maher et al. 2009; Shah et al. 2009; Berger et al. 2010; Tuch et al. 2010; Wang et al. 2012). Analysis of chromatin modification is in its infancy as applied to the cancer cell, but the recent reporting of the ENCODE Project Consortium's genome-wide results (The ENCODE Project Consortium 2012) may provide the tools and technologies to enable new approaches. The technology behind DNA sequencing is improving rapidly in accuracy, cost reduction, and speed, making advances in cancer biology and clinical testing, all based on analysis of the primary sequence of the tumor genome, an essential strategy in the war on cancer. However, the coordinated acquisition and integrated interpretation of all this data has been possible because of a reference genome for comparison.

What have we learned so far?

**Mutation frequencies**

By patient

The median frequency of point mutation varies over more than three orders of magnitude across human tumors; within a given tumor type, the variation in frequency is about one order of magnitude (Fig. 2A).
Inférence

En général, on veut séquencer le cancer et le génome normal pour comparaison

difficultés :

★ *matched normal* contient des cellules somatiques (→ mutations dans la lignée germinale ?)
★ échantillon de tumeur est mixte (normal+lignées clonales)
★ loss-of-heterozygosity
★ erreurs du mappeur

LOH

= loss-of-heterozygosity

★ inférence par probabilité postérieure d’état joint

★ ignore impureté

★ ignore QUAL

★ ignore MAPQ
bcp de différence entre les outils

Fig. 3. Overlaps between somatic SNV candidate sets in the filtered output for the CML exome

Roberts & al Bioinformatics 29 :2223 (2013)
Figure 4 | Diversity and frequency of genetic changes leading to deregulation of signalling pathways in CRC. Non-hypermethylated (nHM; n = 165) and hypermethylated (HM; n = 30) samples with complete data were analysed separately. Alterations are defined by somatic mutations, homozygous deletions, high-level focal amplifications, and, in some cases, by significant up- or downregulation of gene expression (IGF2, FZD10, SMAD4). Alteration frequencies are expressed as a percentage of all cases. Red denotes activated genes and blue denotes inactivated genes. Bottom panel shows for each sample if at least one gene in each of the five pathways described in this figure is altered.
Réarrangements génomiques

breakage/fusion/bridge

B/F/B cycles can loop out to form DM chromosomes.\((34,40)\) The amplified regions can then be transferred to other chromosomes, since DM chromosomes can reintegrate at other locations.\((43,104,105)\) Amplified regions can also be transferred to other chromosomes following fusion of the chromosome containing the amplified DNA with other chromosomes (Fig. 2).\((21)\) As a result, many amplification events resulting from B/F/B cycles in cancer cells will not be located on the end of the chromosome on which they originated, and therefore will not be recognizable as having originated through B/F/B cycles. The initiation of gene amplification by B/F/B cycles would explain why inverted repeats are commonly observed in amplified regions in human cancer cells.\((106)\)

A fourth type of rearrangement caused by telomere loss is translocation, which can be either duplicative or nonreciprocal, both of which are commonly associated with human cancer.\((73,86)\) The analysis of cells actively undergoing B/F/B cycles demonstrates that these translocations are one of the most common mechanisms for telomere addition during B/F/B cycles (unpublished observation). The presence of short terminal deletions (unpublished observation) can also lead to the formation of DM chromosomes that can be involved in high-copy gene amplification or reintegrate back into other chromosomes.

End-sequence profiling

1. séquencer les extrémités des chormosomes artificiels du génome de tumeur

2. mapping

3. inférer les réarrangements

problème algorithmique : scénario de réarrangements pour transformer la référence dans le génome du tumeur

Inférence de réarrangements

(bleu, rouge : réarrangements inter- et intra-chromosomaux ; vert, bleu : réarrangements entre régions répétées ou non)
Évolution du cancer

comment inférer ?
1. échantillons en temps et espace
2. échantillons de cellules singulaires

Exemple : single cells

1. microdissection de cellules isolées
2. clonage + génotypage (120 loci)
3. arbre phylogénétique (neighbor-joining)

Exemple : régions multiples

1. échantillons de régions multiples
2. séquençage complet de l’exome
3. analyse phylogénétique

Régions multiples

(bleu=absent ; gris=présent)

However, the relapse clone evolved from clone 4. A single clone cells were from clone 4, indicating that it may have arisen last (Fig. 2a). Only 5.10% of the tumour growth or survival advantage, as 53.12% and 29.04% of the tumour
1. The additional mutations in clones 2 and 3 may have provided a chemotherapy resistance; although none had translational consequences, we cannot rule out a relevant regulatory mutation in this cluster. We do not know whether any of the cluster 4 mutations conferred chemotherapy and evolved to become the dominant clone at relapse. It is likely that a single cell from clone 3 derived from clone 1, because virtually all the cells in the sample contain the cluster 1 mutations (Fig. 2a). It is possible that a single cell from clone 3 to be present in virtually all the tumour cells at presentation and at relapse, as the variant frequency of these mutations is probably relevant for pathogenesis; one subclone within the founding clone probably contain only this set of mutations. Clones 2, 3 and 4 evolved from clone 1. Clone 1 is the founding clone; 12.74% of the tumour cells
for the primary bone marrow sample, see Supplementary Informa-
primary tumour sample (with a malignant cellular content at 93.72%

(8 patients avec acute myeloid leukaemia ; séquençage de génomes complets)

Histoires différentes

l’histoire peut être linéaire, ou divergente ; on a souvent de la convergence (effet de sélection similaire)

---

Panel with permission, from

RUNX1

MLL3

PIK3CA

thicknesses of the branches reflect the numbers of mutations within each distinct mutation 'cluster'. This gives an indication

frequencies (corrected for local copy number). In the tree in panel
delineation of distinct clusters of mutations — these clusters consist of groups of mutations that share similar mutant allele

populations, and the numbers represent the number of copies of each adjacent gene. Solid lines represent the most likely
evolution, in which the same genetic consequence independently emerges in separate clades of the phylogenetic tree
evolution, in which the same genetic consequence independently emerges in separate clades of the phylogenetic tree

evolution, in which the same genetic consequence independently emerges in separate clades of the phylogenetic tree

identifies the post

range of cancers. Linear evolution (panel
range of cancers. Linear evolution (panel
range of cancers. Linear evolution (panel

and their genetic relationships. For example, both linear and

and their genetic relationships. For example, both linear and

and their genetic relationships. For example, both linear and

probably invalid (as discussed in the text), and molecular

were constant, then this would correlate with chronological

so-called 'molecular clock'. If mutation rates per unit time

the number of mutations that occurs in that lineage: a

the number of mutations that occurs in that lineage: a

the number of mutations that occurs in that lineage: a

mutations that occur after the most recent appearance of a

final complete selective sweep within the cancer: all clonal

clonal mix.

information about the genetic diversity of a cancer and

biology and sampling techniques coupled with existing and

tumour is inferred to have evolved. As discussed in the text,

A phylogenetic tree is a pictorial representation of how a

A phylogenetic tree is a pictorial representation of how a

A phylogenetic tree is a pictorial representation of how a

Box 1 |

The branching structure of the tree captures the number

...
The future is now

★ milliers de génomes ⇒ classification et diagnostique
★ séquençage de régions multiples, à temps multiples ⇒ modélisation de l’écologie du tumeur
★ séquençage de cellules singulaires