Estimating crypt fission rate using approximate Bayesian computation

To estimate the crypt fission rate of the normal colon we used approximate Bayesian computation (ABC). This involved simulating each biopsy, that is each independent sample of colon, many times assuming different values of the crypt fission rate parameter. Each simulation yielded a vector of summary statistics that were compared with the same summary statistics calculated for the observed data. Crypt fission rates for simulations that produce summary statistics similar to those calculated for the observed data are more likely to reflect the true crypt fission rate. This allowed us to approximate the posterior distribution of this parameter without knowing the likelihood function.

Our analysis made the following three assumptions:

1. The crypt fission rate is constant after 4 years of life and is the same in all sectors of the colon and the ileum.
2. The number of crypts is constant after 4 years of life.
3. The mutation burden of substitution signature 1 in each sector of the colon is fixed through life.

We used as summary statistics a vector containing the number of coalescent events observed in 10-year intervals across the entire cohort except we ignored events occurring earlier than 4 years in molecular time to avoid counting events occurring as part of neonatal expansion of the colon.

To time the coalescent events in the observed data, we used the number of mutations on each edge of the phylogenic trees that were assigned to substitution signature 1. Signature 1 represents a clock-like mutational process (<https://www.nature.com/articles/ng.3441>) and under assumption 3 above, the timing of the coalescent events could be estimated given the location of the biopsy and the mutation rate in different sectors of the colon (16.8, 16.1, 12.8 and 12.7 mutations per year in the right, transverse and left side of the colon and the ileum, respectively, see main text).

We generated spatial relationship matrices for each biopsy by reviewing the images from the laser capture microscope. Sometimes, there existed within a site groups of crypts where the distances between crypts were known but distances between groups could not be established. These were treated as independent biopsies in the simulations and only coalescent events between crypts from the same group were counted in the observed data. We were able to quantify the distance between 459 pairs of crypts. A total of 324 crypts across 102 independent biopsies were used in the simulations.

We drew crypt fission rates from a uniform prior between zero fission rate to 1 fission per crypt every 4 years. Each biopsy was simulated as a two dimensional n x n grid of cells with each cell representing a crypt and where n equals three times the largest distance seen between any two crypts from that biopsy. We drew the time until the next crypt fission event from exponential distribution. At each event, a cell in the grid was randomly chosen to die and be replaced by one of its neighbours, chosen at random. This process was repeated until the total time passed exceeded the age of the patient from which the biopsy was drawn. We then sampled the grid in a way that preserved the spatial relationships between the crypts in the observed data and identified the timing of the coalescent events linking the sampled cells.

To estimate the posterior distribution, we used the neural network regression algorithm implemented in the abc package in R (<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.2041-210X.2011.00179.x>, <http://membres-timc.imag.fr/Olivier.Francois/blum_francois09.pdf>).