Spatial heterogeneity of KRAS mutations in colorectal cancers in northern France

**Background:** Somatic mutations in the *KRAS* gene are the most common oncogenic mutations found in human cancers. However, no clinical features have been linked to KRAS mutations in colorectal cancer (CRC).

**Purpose:** In this study, we attempted to identify the potential geographical population clusters of *KRAS* mutations in CRC patients in northern France.

**Patients and methods:** All patients with CRC who were identified to have *KRAS* mutations between 2008 and 2014 at the Regional Molecular Biology Platform at Lille University Hospital were included. 2,486 patients underwent a *KRAS* status available, with 40.9% of CRC with KRAS mutations in northern France. We retrospectively collected demographic and geographic data from these patients. The proportions of *KRAS* mutation were smoothed to take into account the variability related to low frequencies and spatial autocorrelation. Geographical clusters were searched using spatial scan statistical models.

**Results:** A mutation at *KRAS* codon 12 or 13 was found in 1,018 patients [40.9%]. We report 5 clusters of over-incidence but only one elongated cluster that was statistically significant [Cluster 1; proportion of *KRAS* mutation among CRC: 0.4570; RR=1.29; P=0.0314]. We made an ecological study which did not highlight a significant association between *KRAS* mutations and the distance to the Closest Waste Incineration Plant, and between *KRAS* mutations and The French Ecological Deprivation Index but few socio-economic and environmental data were available.

**Conclusion:** There was a spatial heterogeneity and a greater frequency of *KRAS* mutations in some areas close to major highways and big cities in northern France. These data demand deeper epidemiological investigations to identify environmental factors such as air pollution as key factors in the occurrence of *KRAS* mutations.

**Keywords:** colorectal cancer, *KRAS* mutation, spatial scan statistics, carcinogens, highways

---

**Novelty and impact**

Our results suggest a spatial heterogeneity of KRAS mutations in colorectal cancer across northern France. We identified 5 clusters with a high proportion of *KRAS* codon 12-13 mutations, and only one elongated cluster close to major regional highways was statistically significant. Further studies are warranted to elucidate the causal role of carcinogens in the occurrence of *KRAS* mutations.

**Introduction**

*KRAS* gene mutations are the most common oncogenic mutations in human cancers.

They are heterozygous mutations corresponding to a substitution of a single amino acid which prevents the dephosphorylation and thus activates the KRAS protein, an oncogenic GTPase.
There is a strong correlation of KRAS genotype (i.e., mutated or not) with primary cancer and metastasis in metastatic colorectal cancer (mCRC). This underlines the early characteristics of KRAS mutation, which essentially drives the adenoma-cancer disease progression culminating into metastasis. In 2008, KRAS mutational status was associated with good outcomes in mCRC patients treated with cetuximab and panitumumab. Furthermore, these studies showed that in patients with mutated KRAS gene, cetuximab had little or no effect compared to standard chemotherapy. The results implied that EGFR inhibitors would be effective only in cases with wild-type RAS tumor including NRAS and/or KRAS mutations, representing approximately 50% of patients. The systematic investigation of KRAS and NRAS mutational status for exons 2, 3, and 4 was thus required before any prescription of anti-EGFR.

The KRAS mutation rate appears to differ based on geographic location. In Asia, 33.3% of CRC patients have been identified to have KRAS mutations. In Europe, the proportion of CRC patients with KRAS mutations is 36%, with 38% in France. The spatial heterogeneity of CRC has been demonstrated at the regional and cantonal levels in France. No study has yet clarified this result by characterizing the spatial distribution of the genotypes.

The aim of this study was to analyze the spatial variations of the KRAS mutation proportion in CRC, in northern France on a small geographical scale, during the period from 2008 to 2014. Furthermore, the link between the spatial variations of KRAS mutation proportion and the distance to closest waste incineration plant (CWIP) and the social deprivation was investigated by means of an ecological study in the goal to identify environmental factors as putative key factors in the occurrence of KRAS mutations.

Materials and methods
This study was conducted in northern France, which had a total population of 4,050,756 inhabitants at the time of the 2012 national census. This area is divided into 170 cantons [a French small administrative area with population ranging from 12,934 to 231,491 inhabitants]. The centroid of each canton as defined by the geographical center (latitude and longitude) was used for the spatial analyses.

Source of epidemiological data
All patients with CRC who underwent KRAS mutation status determination between 2008 and 2014 at the Regional Molecular Biology Platform at Lille University Hospital were included in our cohort. The retrospective data collected were demographic data: sex, age, canton of residence at the time of primary diagnosis; the KRAS mutational status: absence/presence of mutation and type. The data were deidentified.

This study was approved by the CCTIRS (Comité Consultatif sur le Traitement de l’Information en matière de Recherche dans le domaine de la Santé, approval n° 14.728) and is in full accordance with French laws.

KRAS mutation analysis
Genomic DNA was isolated from 10-µm paraffin-embedded sections using the QIAamp DNA FFPE Tissue DNA extraction kit (Qiagen, Courtaboeuf, France) according to the manufacturer’s protocol. Mutation analysis of KRAS codons 12 and 13 was performed by pyrosequencing using the PyroMark™ MD system (Biotage, Qiagen). Genomic DNA (25 ng) was amplified by PCR and sequenced with the PyroMark™ KRAS kit according to the manufacturer’s instructions (Qiagen). DNA samples from cell lines known to be KRAS mutated and KRAS wild-type were introduced in each series as positive and negative controls, respectively. Each detected mutation was checked by a new amplification and analysis. The lower limit of detection of pyrosequencing assay was between 5% and 10% mutated DNA.

Ecological data
Deprivation index
The French Ecological Deprivation Index (EDI) was computed for each canton for measuring the deprivation. This index is composed of 10 variables: overcrowding, no access to a system of central or electric heating, non-owner, unemployment, foreign nationality, no access to a car, unskilled worker-farm worker, household with more than 6 persons, low-level of education and single-parent household. Higher index in a canton correlated with higher deprivation. Socioeconomic data were extracted from the 2012 national population census, provided by the French national institute Institut National de la statistique et des études économiques. The obtained French EDI was considered as quartiles for ecological regression described in the section “Statistical analyses.”

Distance to closest waste incineration plant (CWIP)
The data of the French national institute Agence de l’Environnement et de la Maîtrise de l’Energie were used.
to identify the 10 waste incineration plants in activity before and during the period of the study and located in the studied geographical area. For each canton, the Euclidian distance in kilometers between its centroid and the CWIP were computed. The obtained distance to CWIP was considered as quartiles for ecological regression.

Statistical analyses
Descriptive analysis
Quantitative variables were described using mean (SD) in case of normal distribution or using median (interquartile interval) otherwise. The normality of the distributions was assessed using histograms, normal probability plot and the Shapiro–Wilk’s test. Qualitative variables were described by means of percentages (frequencies).

Smoothing the proportion of KRAS mutations
The crude proportion of KRAS mutations by cantons showed high instability due to low frequencies. A spatial smoothing method was used to reduce this instability and take into account the spatial autocorrelation between neighboring cantons. We used a hierarchical Bayesian three-level binomial model as proposed by Besag et al.\(^\text{15}\) The number of CRC with KRAS mutation per canton was considered as an outcome variable and the total number of CRC per canton as an offset. In this mixed model, the variances of the two random spatial effects have a non-informative gamma prior distribution as suggested by Bernardinelli et al.\(^\text{16}\) The model was fitted using the Integrated Nested Laplace Approximation (INLA) method.\(^\text{17,18}\) A 95% Bayesian credible interval (BCI) was associated with each proportion estimate.

Spatial cluster detection
The detection of significant spatial clusters of cantons with a high proportion of KRAS mutation was performed. We used spatial scan statistics on the basis of a Bernoulli model to test the presence of KRAS mutation clusters and identify their location without pre-selection bias.\(^\text{19}\) These methods are powerful and largely used in the field of epidemiology.\(^\text{20–22}\) We used an elliptic spatial scan statistic that allows the detection of cluster of an elliptical shape, including circular shape than can be considered as a special case of an ellipse.\(^\text{23}\)

We observed that circular clusters may still be highlighted since the circle can be considered as a special case of an ellipse. The significance of each detected clusters has been evaluated using 9,999 Monte-Carlo replications under the null hypothesis of absence of cluster. A RR was associated with each significant cluster. The RR was interpreted as the risk of observing a KRAS mutation in the cluster compared to the rest of studied area.

Ecological regression
In order to analyze the association between geographical variations of KRAS mutation proportion and the ecological variables, we used an extension of the previous hierarchical Bayesian model that includes the two ecological variables (French EDI and distance to the CWIP) as fixed effects in the model with a non-informative prior as suggested by Lawson.\(^\text{24}\) First quartile of each ecological variable was considered as the reference. Each quartile was assigned an RR and its BCI at the 95% level.

Smoothing of KRAS mutation proportions and ecological regression analyses were carried out using the R software v. 3.4.3 (package: R-INLA).\(^\text{25}\) The cluster detection was performed using the SaTScan software v. 9.6.\(^\text{26}\) Maps were performed using QGIS software v. 2.18.\(^\text{27}\) All statistical analyses were considered significant at 0.05 type 1 error.

Results
During the period from 2008 to 2014, 2486 patients underwent a KRAS mutation status determination in northern France. The median age was 61 years (54;71) and 41.5% of the patients were women. Among the 2486 patients, 1018 showed a KRAS mutation (40.9%). The most frequent mutations observed were Gly12Asp (14.3%, n=355), Gly12Val (9.8%, n=243) and Gly13Asp (6.8%, n=169) (Figure 1).

Spatial analysis
Among the 170 cantons of the studied area, the median number of CRC per canton was 9 (4; 20) and the median number of CRC with KRAS mutation per canton was 3 (1; 8). Only three cantons did not have CRC cases.

Smoothed proportions of KRAS mutation
The smoothed proportions of KRAS mutation varied from 0.2829 (95% BCI: [0.1804; 0.3941]) in the canton of Dunkerque, located in the north of the studied area, to 0.5126 (95% BCI: [0.3163; 0.7187]) in the canton of Outreau, close to the city of Boulogne-sur-Mer. The average-smoothed proportion of KRAS mutation was 0.3952 (95% BCI: [0.2606; 0.5418]). The geographical distribution of the smoothed proportions of KRAS mutation is shown in Figure 2.
Spatial clusters detection
The elliptic spatial scan statistics detected 5 clusters of high proportion of KRAS mutation. The clusters characteristics are presented in Table 1. Only one elongated cluster was statistically significant ranging from the city of Tourcoing to the city of Marquion (Cluster 1; smoothed proportion of KRAS mutation: 0.4570; RR=1.29; P=0.0314) close to the major highways. This cluster is mapped in Figure 3.

Ecological regression
Using the first quartile as reference, the RR of KRAS mutation was assessed for each level of distance of the CWIP and French EDI (Figure 4). Based on the 95% BCI associated to the RRs, no significant association was highlighted between KRAS mutation and the distance to the CWIP (Q2: RR=0.82, 95% BCI [0.60;1.12]; Q3: RR=0.93, 95% BCI [0.67;1.30], Q4: RR=0.98, 95% BCI [0.64;1.49]). According to the same methodology,
Table 1 Spatial clusters of high proportion of KRAS mutations identified by the elliptic spatial scan statistics in northern France, 2008–2014

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of cantons</th>
<th>Populationa</th>
<th>Observedb</th>
<th>Expectedc</th>
<th>Proportiond</th>
<th>Smoothed proportione</th>
<th>Pr</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>509</td>
<td>254</td>
<td>208.79</td>
<td>0.4990</td>
<td>0.4570</td>
<td>0.0314</td>
<td>1.29</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>2.46</td>
<td>1.0000</td>
<td>0.4089</td>
<td>0.7949</td>
<td>2.45</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>70</td>
<td>42</td>
<td>28.71</td>
<td>0.6000</td>
<td>0.4523</td>
<td>0.8253</td>
<td>1.48</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>43</td>
<td>28</td>
<td>17.64</td>
<td>0.6512</td>
<td>0.4074</td>
<td>0.8390</td>
<td>1.6</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>3.69</td>
<td>0.8889</td>
<td>0.5126</td>
<td>0.9346</td>
<td>2.18</td>
</tr>
</tbody>
</table>

Notes: aNumber of patients inside the clusters; bobserved number of patients with KRAS mutation; cexpected number of patients with KRAS mutation under the hypothesis of spatial homogeneity; dcrude proportion of KRAS mutation; emedian of the smoothed proportions inside the cluster.

Figure 3 Spatial cluster of high proportion of KRAS mutations identified by elliptic spatial scan statistics, northern France, 2008–2014.

Figure 4 RRs (95% BCI) of KRAS mutation by French EDI and distance (in km) to the CWIP quartiles.

Abbreviations: BCI, Bayesian credible interval; EDI, Ecological Deprivation Index; CWIP, closest waste incineration plant.
no significant association between KRAS mutation and deprivation index was found (Q2: RR=0.98, 95% BCI: [0.70;1.37], Q3: RR=1.07, 95% BCI: [0.77;1.48], Q4: RR=0.93, 95% BCI: [0.69;1.25]).

Discussion

We highlighted significant spatial variations in the proportion of KRAS mutations in northern France. In addition, the implementation of elliptic spatial scan statistic made it possible to reveal the existence of a significant spatial cluster of a higher proportion of KRAS mutations.

The proximity of the significant cluster with regional highways nodes is questioning (Figure 3), requiring further ecological investigations.

In the literature, only two studies proposed a geographic but not spatial analysis of the rate of KRAS mutated CRCs: in these studies, the country was divided into 4 or 5 large regions and inter-regional comparisons of the proportion of KRAS mutated cancers were performed. In India, a cohort-study of 1323 patients revealed that the KRAS mutations were more common in northern India than in the rest of the country (in 339 patients, 86 KRAS mutant versus 253 KRAS wild type; \(P<0.01\)). In Brazil, of the 8234 patients analyzed, KRAS mutations were more frequent in parts of Southeast and North. The geographic distribution of KRAS mutations was not similar, for example, the mutation Gly12Asp (GGT > GAT) was the highest in most regions with the exception of the North where Gly12Val (GGT > GTT) was the most frequent. Thus, the authors hypothesized that this observation could be explained by the large ethnic disparity in Brazil with a prevalence of migrant European settlers in the southern region and a prevalence of American native people in the north.

In USA, KRAS mutations were statistically less frequent in non-Hispanic Whites when compared to that in African-Americans (\(P=0.0001\)). A possible explanation of the different rates of KRAS mutations observed in the populations would be the disparity in testing due to different economic status. Another possible explanation for the geographic disparity of KRAS mutation could be lifestyle.

Martinez et al found a correlation between tobacco consumption and the presence of KRAS mutations in colorectal polyps and the same correlation was reported by Curtin et al in rectal cancers. Similarly, alcohol was also associated with increased RR of colon cancer with KRAS mutation. A significant relationship was also established with obesity, particularly with the waist–hip ratio in women and KRAS-mutated CRCs. The use of alcohol and tobacco in northern France, which is higher than in other French regions, and the high prevalence of obesity could be confounding factors in the geographic clustering observed.

Our study has some limitations. There was in northern France between 2008 and 2010 an estimate of 1283 (1196–1376) incident-cases per year. Over a 6-year period, our study may have only included less than one-third of the total CRC patients with KRAS mutations. The recruitment of KRAS status via a reference center for a test routinely prescribed by the other centers suggests the representativeness of the study population compared to the overall regional population. Nevertheless, if in some centers KRAS mutation determination was more frequently requested in non-metastatic patients, the prevalence could have been increased superficially resulting in a greater number of patients to be included in our study. Another bias could be relocation if a significant number of patients migrated from another region since, at the time of diagnosis, we recorded the place of residence and not the geographical history of patients.

In the ecological study, no significant association was highlighted between KRAS mutations and the distance to the CWIP and between KRAS mutations and deprivation index but few socio-economic and environmental data were available.

Outdoor air pollution, known as carcinogen (Group 1 in the IARC classification), is a mixture of multiple pollutants. Although this classification is for lung cancer, it is relevant to question the impact of such pollution on CRC. Specifically, some traffic-related pollutants, such as NO\(_2\) and benzene, are suspected to be involved in the pathogenesis of CRC. High-level exposure to diesel emissions may increase the risk of rectal cancer in Canadian men, not colonic cancer. A study including more than 600,000 adults in the US who participated in the Cancer Prevention Study II and who were followed for 22 years (from 1982 to 2004), examined associations of mortality from cancer at 29 sites with long-term ambient exposure to three ambient pollutants: PM\(_{2.5}\), nitrogen dioxide (NO\(_2\)) and ozone (O\(_3\)). In over 43,000 non-lung cancer deaths, exposure to NO\(_2\), a pollutant considered consensually to be a marker of traffic pollution, was associated with CRC death, with a 6% increase per 6.5 ppb increment. The presence of the cluster along the major axes could also be explained by the presence of polluting industries, located close to the motorways. Industrial emissions can
contribute to cumulative and long-term exposure of populations. An increase in mortality due to CRC was linked to the vicinity of the polluting industries (i.e., metal production, paper and wood production) in Spain.\textsuperscript{37}

Our results are inconsistent with those of a study that suggests pesticide exposure as another explanatory factor.\textsuperscript{38} Agricultural areas with high pesticide use are clustered in the North and Southeastern regions of our study area. However, these areas did not feature among regions with high mutation proportions or in the cluster.

These data focused on CRC highlighted the current lack of information on the relationship between environmental pollution and KRAS mutations. Deepening this link will improve understanding of the underlying mechanisms of mCRC.

Conclusion
There is a spatial heterogeneity suggesting a higher prevalence of KRAS mutated CRC in geographical areas in Northern France. One relevant cluster was in close vicinity to highways network. Epidemiological investigations should be conducted to identify environmental factors in the occurrence of KRAS mutations.

Highlights
- Our results suggest a spatial heterogeneity of KRAS mutations in colorectal cancer across northern France.
- We identified 5 clusters with a high proportion of KRAS codon 12–13 mutations.
- Only one elongated cluster including a major highway was significant.

Acknowledgment
We would like to thank Severine Marchant (Centre Oscar Lambret) for editing the manuscript and Editage (www.editage.com) for English language editing. The abstract has been accepted for the ASCO 2019 annual meeting in Chicago, IL; abstract n°e15101, citation: J Clin Oncol 37, 2019 (suppl; abstr e15101).

Disclosure
A travel grant by Sanofi-Genzyme was obtained by Anthony Turpin. Dr Anthony Turpin reports personal fees from Amgen, Merck, Servier, and grants from Astra-Zeneca, Pfizer, outside of the submitted work. Dr Clotilde Descarpentries reports non-financial support from Astra-Zeneca and Roche, outside the submitted work. The authors report no other conflicts of interest in this work.

References