



UNIVERSITY OF TAMPERE

This document has been downloaded from
TamPub – The Institutional Repository of University of Tampere

 *Publisher's version*

The permanent address of the publication is <http://urn.fi/URN:NBN:fi:uta-201309201355>

Author(s): Kaasinen, Eevi; Aavikko, Mervi; Vahteristo, Pia; Patama, Toni; Li, Yilong; Saarinen, Silva; Kilpivaara, Outi; Pitkänen, Esa; Knekt, Paul; Laaksonen, Maarit; Artama, Miia; Lehtonen, Rainer; Aaltonen, Lauri; Pukkala, Eero

Title: Nationwide registry-based analysis of cancer clustering detects strong familial occurrence of Kaposi sarcoma

Year: 2013

Journal Title: Plos ONE

Vol and number: 8 : 1

Pages: 1-7

ISSN: 1932-6203

Discipline: Biomedicine; Cancers; Health care science

School /Other Unit: School of Health Sciences

Item Type: Journal Article

Language: en

DOI: <http://dx.doi.org/10.1371/journal.pone.0055209>

URN: URN:NBN:fi:uta-201309201355

URL: <http://dx.doi.org/10.1371/journal.pone.0055209>

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.

Nationwide Registry-Based Analysis of Cancer Clustering Detects Strong Familial Occurrence of Kaposi Sarcoma

Eevi Kaasinen^{1,2,9}, Mervi Aavikko^{1,2,9}, Pia Vahteristo^{1,2}, Toni Patama³, Yilong Li^{1,2}, Silva Saarinen^{1,2}, Outi Kilpivaara^{1,2}, Esa Pitkänen^{1,2}, Paul Knekt⁴, Maarit Laaksonen⁴, Miia Artama³, Rainer Lehtonen^{1,2}, Lauri A. Aaltonen^{1,2*}, Eero Pukkala^{3,5}

1 Genome-Scale Biology Research Program, and Department of Medical Genetics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, **2** Haartman Institute, University of Helsinki, Helsinki, Finland, **3** Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland, **4** National Institute of Health and Welfare, Helsinki, Finland, **5** School of Health Sciences, University of Tampere, Tampere, Finland

Abstract

Many cancer predisposition syndromes are rare or have incomplete penetrance, and traditional epidemiological tools are not well suited for their detection. Here we have used an approach that employs the entire population based data in the Finnish Cancer Registry (FCR) for analyzing familial aggregation of all types of cancer, in order to find evidence for previously unrecognized cancer susceptibility conditions. We performed a systematic clustering of 878,593 patients in FCR based on family name at birth, municipality of birth, and tumor type, diagnosed between years 1952 and 2011. We also estimated the familial occurrence of the tumor types using cluster score that reflects the proportion of patients belonging to the most significant clusters compared to all patients in Finland. The clustering effort identified 25,910 birth name-municipality based clusters representing 183 different tumor types characterized by topography and morphology. We produced information about familial occurrence of hundreds of tumor types, and many of the tumor types with high cluster score represented known cancer syndromes. Unexpectedly, Kaposi sarcoma (KS) also produced a very high score (cluster score 1.91, p -value <0.0001). We verified from population records that many of the KS patients forming the clusters were indeed close relatives, and identified one family with five affected individuals in two generations and several families with two first degree relatives. Our approach is unique in enabling systematic examination of a national epidemiological database to derive evidence of aberrant familial aggregation of all tumor types, both common and rare. It allowed effortless identification of families displaying features of both known as well as potentially novel cancer predisposition conditions, including striking familial aggregation of KS. Further work with high-throughput methods should elucidate the molecular basis of the potentially novel predisposition conditions found in this study.

Citation: Kaasinen E, Aavikko M, Vahteristo P, Patama T, Li Y, et al. (2013) Nationwide Registry-Based Analysis of Cancer Clustering Detects Strong Familial Occurrence of Kaposi Sarcoma. PLoS ONE 8(1): e55209. doi:10.1371/journal.pone.0055209

Editor: Olga Y. Gorlova, The University of Texas M. D. Anderson Cancer Center, United States of America

Received: September 12, 2012; **Accepted:** December 23, 2012; **Published:** January 24, 2013

Copyright: © 2013 Kaasinen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Academy of Finland Center of Excellence in Cancer Genetics Research (No 250345) and Postdoctoral Researcher's grants (No 21290 and 137680), European Research Council (No 268648 - NGG), Cancer Society of Finland, Ida Montini Foundation, Paulo Foundation, Orion Pharmos Foundation, and Sigrid Juselius Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lauri.aaltonen@helsinki.fi

9 These authors contributed equally to this work.

Introduction

Identification of inherited genetic factors that contribute to the development of cancer has been a subject of extensive research since the discovery of the first tumor susceptibility genes. Various approaches have been used to estimate the proportion of the inherited versus environmental factors in the etiology of different cancers. Twin studies have elucidated a large heritable fraction (26%–42%) for many common cancers, such as those of stomach, colon/rectum, breast, prostate, and lung [1]. Systematic registry-based analyses have shown a significantly increased risk for example for brain tumors, various lymphoma subtypes, and gastrointestinal neuroendocrine tumors in family members of affected individuals [2–5].

The earliest examples of cancer predisposing genes were identified in large cancer syndrome families. These families are rare and the contribution of these genes to the cancer burden as

a whole is relatively small. For example, high-penetrance mutations in genes such as *MLH1*, *MSH2*, *APC*, and *MYH* account for only about 5% of all colorectal cancer cases [6]. However, identification of cancer predisposing genes has contributed significantly to our understanding of cancer biology, since they are often central also in the development of sporadic tumors.

The genetic background of the Finnish population (currently 5.4 million) is relatively well-studied and, even though considered relatively homogenous, has been shown to harbor substantial substructuring [7]. Ten distinct subpopulations of Finns have been characterized with high-density SNP genotyping suggesting multiple bottlenecks in the history of the Finnish population. After internal migration in the 1500s, population size of the wide inland areas of Finland grew in isolation because of long distances, and small villages formed genetically distinct subpopulations. Some degree of genetic admixture and migration due to

urbanization started taking place after World War II. Single-origin mutations which have spread without negative effect on biological fitness have accounted for highly prevalent diseases in some of these isolated populations. This founder effect has enabled successful identification of genetic susceptibility to diseases in Finland (reviewed in [8]). As an example, the discovery of familial pituitary adenoma clusters in Northern Finland led to the identification of *aryl hydrocarbon receptor interacting protein (AIP)* defects in patients with pituitary adenoma predisposition [9]. Mutations in *AIP* originally identified in Finland have subsequently been found from similar patients worldwide [10].

The Finnish Cancer Registry (FCR) is a nation-wide database covering currently more than one million patients, and practically all incident cancers and cancer deaths in Finland since 1953 [11]. Physicians, hospitals, and pathology and hematology laboratories send cancer notifications to FCR. This reporting was made obligatory by law in 1961. Thus, FCR has also a link to sample information, such as archival tissue blocks, for all the patients that have been examined in pathology laboratories. Information about family history and family relationships is not recorded in the cancer notifications, but identification of patients' relatives is possible through the National Population Registry (NPR) and the Parish Registries (PRs). PRs have kept records of citizens since 1580 until NPR was established. The personal identity code (PIC) system was implemented in Finland in the 1960s, and PIC is used as the identification key also in FCR for all patients except for those who died before 1967. Both FCR and NPR have an excellent coverage and high quality data [11,12].

Conventional epidemiological tools are poorly suited to search for evidence of cancer predisposition, in particular if the conditions in question are uncommon. Here we describe an effective method that utilizes the extensive patient data in FCR, and information about family name at birth and municipality of birth in NPR, to systematically identify familial aggregation of all different tumor types on a nation-wide scale. The method enables identification of families suitable for sample collection and further research of the underlying genetic predisposition. Similar approach in a smaller scale has earlier been exploited in a search for prostate cancer families in Finland [13]. In this study, we apply our method to cluster all available patients in FCR. We calculate statistical significance for each cluster, and produce information about familial occurrence of hundreds of tumor types. We describe in detail a strong familial aggregation and uneven geographical distribution of Kaposi sarcoma (KS) in Finland. Furthermore, we report several families with two first degree relatives and a striking family of five cases with KS.

Materials and Methods

Ethical Considerations

The research registry and data processing procedures were evaluated and approved by the Ministry of Social Affairs and Health in Finland.

Patient Data in the Finnish Cancer Registry

We considered 1,175,040 neoplasms registered in 1953–2011 in FCR. This included 212,685 cases of certain commonly registered precancerous lesions such as basal cell carcinoma of skin and polycythemia vera (registered since 1969). Currently, the disease classification in FCR is based on the ICD-oncology, 3rd edition.

Systematic Clustering

Prior to clustering, the tumor types were classified further to topography and morphology groups (Tables S1 and S2). Tumor

types with at least two patients in FCR were included, resulting in 846 combinations of topography and morphology (Table S3). To obtain information on patients' family names at birth and municipalities of birth, we linked the 878,593 patients with PIC to NPR.

For each tumor type, systematic clustering was performed based on municipality of birth and family name at birth (MN-clusters). The observed number of patients (O) with each tumor type in FCR was calculated in a stratum defined by municipality (M), family name (N), sex, and year of birth. The proportion of persons in each M, N, sex, and year of birth category in the entire population of same gender and year of birth was calculated from the same version of NPR database that was linked with FCR patients. When that proportion was multiplied with the total sex and birth-year specific number of cancer cases of the cancer type in question in the NPR-linked subset of FCR, the expected number of cases (E) in each stratum was obtained. The stratum-specific observed and expected numbers were added up over the gender and birth year categories. O/E ratios were calculated for each MN category, and their 95% confidence intervals were defined assuming a Poisson distribution of observed numbers.

Estimation of Familial Occurrence with Cluster Score

To estimate the familial occurrence of various tumor types, cluster score was calculated. For each tumor type i , we selected clusters with the lower limit of confidence interval (CI-low) greater than or equal to α , and calculated the number of patients ($X_{i\alpha}$) in these clusters.

We define the cluster score of tumor type i to be $X_{i\alpha}$ divided by the total number of patients with the same tumor type (Y_i) per 100,000 persons in Finland. For example, cluster score with $\alpha = 10$ for Kaposi sarcoma was calculated as following ($X = 19$ and $Y = 537$ from Table 1):

$$19/537 * 5,400,000/100,000 = 1.91.$$

To estimate whether $X_{i\alpha}/Y_i$ was different than expected ratio $Y_i * X_{\alpha}/Y$ (where X_{α}/Y is the respective ratio over all tumor types in Finland; 0.0029 for $\alpha = 10$), two-sided Poisson test was calculated with 95% confidence level. The p-values were adjusted for multiple test comparisons with false discovery rate method [14].

Confirmation of Kinship in the Clustered Kaposi Sarcoma Cases

Relatedness of patients in KS clusters was confirmed using information from NPR and PRs. NPR includes a link to parents for persons born since 1950s and who had not died before the PIC system was implemented in the 1960s. For the rest of the patients, we utilized PRs to identify first degree relatives and their children. PRs were also utilized for more thorough genealogy research. The relatives of the clustered KS cases were traced back at least three generations. To identify additional affected cases in the families, the cancer histories were checked from FCR for individuals identified through genealogy.

Geographical Illustration of Kaposi Sarcoma in Finland

The O/E ratios for Kaposi sarcoma cases in each municipality of birth were calculated as described previously for MN-clusters, and illustrated using the small-area mapping method developed in FCR [15]. The twenty-one largest cities in Finland, including Vyborg (currently part of Russia), are shown as circles with

Table 1. Tumor types showing the strongest evidence for familial occurrence.

| Topography | Morphology | Cluster score | p-value* | Number of patients in FCR | Number of clusters** | Number of patients in clusters** | Suggestive predisposition gene |
|---|---------------------------------------|---------------|----------|---------------------------|----------------------|----------------------------------|--------------------------------|
| central nervous system | hemangioblastoma | 4.98 | 2.77E-14 | 141 | 5 | 13 | VHL |
| thyroid gland | medullary carcinoma | 3.55 | 7.86E-21 | 335 | 9 | 22 | RET |
| Skin | Kaposi sarcoma | 1.91 | 2.05E-13 | 537 | 9 | 19 | |
| Pancreas | neuroendocrine carcinoma | 1.40 | 2.75E-06 | 386 | 5 | 10 | MEN1, RET |
| Kidney | nephroblastoma | 1.26 | 5.98E-06 | 428 | 5 | 10 | WT1 |
| small intestine | neuroendocrine carcinoma | 0.98 | 1.19E-09 | 1154 | 10 | 21 | |
| hematopoetic and reticuloendothelial system | myeloid leukemia | 0.69 | 2.75E-09 | 2186 | 14 | 28 | |
| vulva and vagina | squamous cell neoplasms and carcinoma | 0.63 | 1.37E-10 | 3092 | 17 | 36 | |
| mesothelium | mesothelioma | 0.62 | 3.49E-06 | 1750 | 10 | 20 | |
| Breast | mucinous and mucinous cystic tumor | 0.57 | 2.45E-07 | 2564 | 13 | 27 | |
| thyroid gland | papillary adenocarcinoma | 0.56 | 2.77E-20 | 7863 | 41 | 82 | |
| central nervous system | neurofibroma | 0.55 | 1.02E-05 | 2062 | 9 | 21 | NF1 |
| central nervous system | neoplasm malignant | 0.51 | 1.89E-05 | 2346 | 11 | 22 | |
| hematopoetic and reticuloendothelial system | polycythemia vera | 0.50 | 6.07E-05 | 2173 | 10 | 20 | |
| hematopoetic and reticuloendothelial system | chronic lymphatic leukemia, b-cell | 0.48 | 3.37E-13 | 7478 | 31 | 66 | |
| Kidney | neoplasm malignant | 0.47 | 1.20E-05 | 3015 | 13 | 26 | |
| lymph node | Hodgkin lymphoma | 0.46 | 2.77E-09 | 5499 | 22 | 47 | |
| esophagus | squamous cell neoplasms and carcinoma | 0.45 | 8.37E-09 | 5514 | 23 | 46 | |
| Prostate | epithelial neoplasm and carcinoma | 0.44 | 2.82E-05 | 3187 | 13 | 26 | |
| Lip | squamous cell neoplasms and carcinoma | 0.35 | 1.04E-05 | 7167 | 22 | 46 | |

*Poisson distribution, two-sided test with 95% confidence level, p-values adjusted with FDR for multiple comparisons** $Chow \geq 1$.
doi:10.1371/journal.pone.0055209.t001

a diameter relative to E, and with the color shading indicating the O/E ratio in that city. The remaining map area illustrates the floating averages of the O/E ratios of the neighboring municipalities, excluding the largest cities. The map was based on borders of Finland before year 1940 because a large proportion of the KS patients were born before that.

Determination of HHV8 Seroprevalence in Finland

To estimate Human herpesvirus 8 (HHV8) infection rate in Finland, serological testing was performed in age and sex matched serum samples of 200 Eastern and 200 Western Finns obtained from the National Public Health Institute's Health 2000 cohort. The serological analyses were done at HUSLAB (Helsinki, Finland) with direct HHV8 IgG antibody Enzyme Linked ImmunoSorbent Assay.

Results

We clustered 878,593 cases in FCR to assess the familial aggregation of various tumor types defined by topography and morphology. The clustering was performed based on tumor type, municipality of birth, and family name at birth (MN-clusters). As a result, we identified 25,910 MN-clusters representing 183 combinations of topography and morphology (see Table S3 for distribution of cases). All the clusters fulfilled the criterion that the number of cases in the cluster was higher than the expected number based on the distribution of family names and birth municipalities in Finland, with $CI\text{-low} \geq 1$.

We also sought to estimate the tumor types that showed strongest evidence for familial occurrence, and for that we calculated cluster score for tumor types that produced significant MN-clusters. The clusters with the strongest confidence, $CI\text{-low} \geq 10$, were considered, encompassing 4.7% of all MN-clusters and representing 169 tumor types (Table S3). This corresponds to 0.3% of the patients in FCR.

Tumor types and respective cluster scores with FDR adjusted p-values < 0.0001 are displayed in Table 1. Cluster scores and p-values for all tumor types are listed in Table S3. Most of the top scoring tumor types indicated rare cancer predisposition syndromes. The most frequently clustered tumor type was hemangioblastoma, a common finding in patients with the Von Hippel-Lindau syndrome (VHL [MIM #193300]) (cluster score 4.98). The second most frequently clustered tumor type was medullary thyroid carcinoma indicative of patients with Multiple Endocrine Neoplasia type 2 (MEN2A [#171400] and MEN2B [#162300]) (cluster score 3.55). Among the most common tumor types showing strong evidence for familial occurrence were papillary thyroid adenocarcinoma (cluster score 0.56), chronic lymphatic leukemia (cluster score 0.48), and squamous cell carcinoma of the lip (cluster score 0.35), all with over 7000 registered patients in FCR (Table 1).

Kaposi Sarcoma

Familial occurrence of KS, the third most frequently clustered tumor type (cluster score 1.91), was further studied through genealogy as an example of the procedure to be followed with promising clusters in the systematic search. Out of the 23 KS patients in the MN-clusters, 16 (70%) were shown to be first-degree relatives within the cluster, and 14 of these belonged to clusters with $CI\text{-low} \geq 10$ (Fig. 1). The genealogy revealed a family of five affected individuals with KS in two generations (Fig. 2). Two family members belonged to the same MN-cluster, and a third sibling and an affected cousin were connected to the family through genealogy work. We were able to confirm the mother's

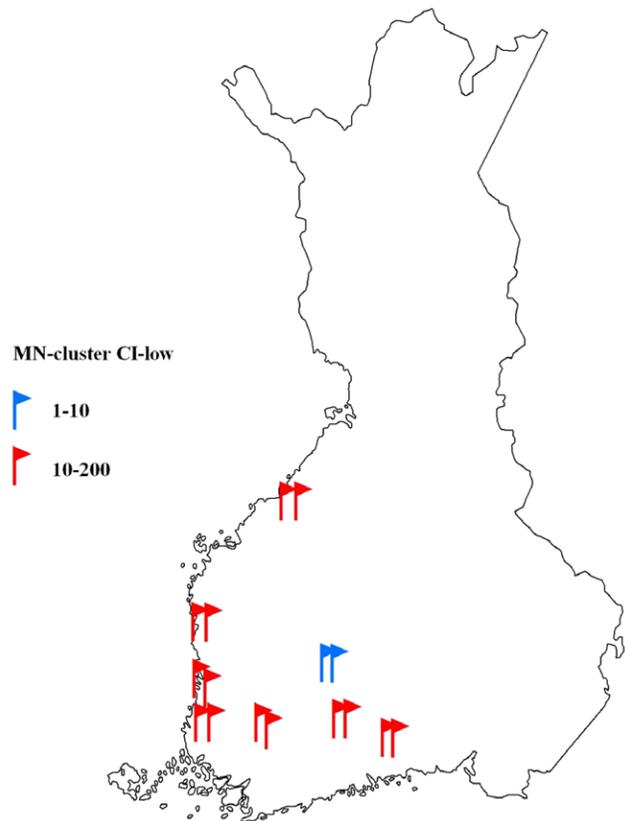


Figure 1. Familial Kaposi sarcoma (KS) cases in Finland; each flag marks a confirmed familial KS patient belonging to a cluster of municipality of birth and family name at birth (MN-cluster). Red and blue colors represent the lower confidence limit (CI-low) of MN-clusters. The borders of Finland are based on those before year 1940.
doi:10.1371/journal.pone.0055209.g001

KS diagnosis (she had died in 1962) from the radiotherapy records after it was verbally reported by her daughter.

The geographical distribution of KS incidence according to municipality of birth of the 537 KS patients in FCR showed a strikingly uneven pattern (Fig. 3). Patients accumulate almost entirely to the rural areas of Western Finland. Another aggrega-

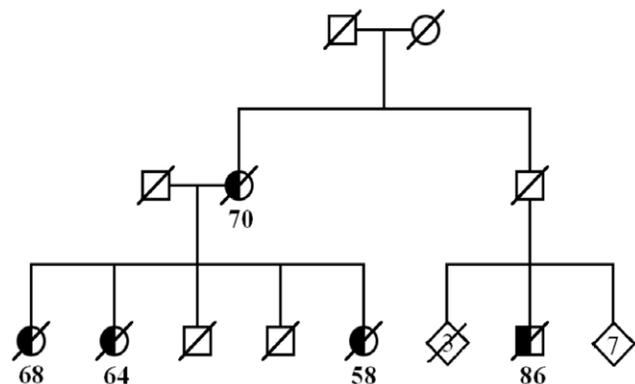


Figure 2. Expansion of a key pedigree; a family of five patients with Kaposi sarcoma (KS) from Eastern Finland. The age at diagnosis of KS is shown for each affected individual.
doi:10.1371/journal.pone.0055209.g002

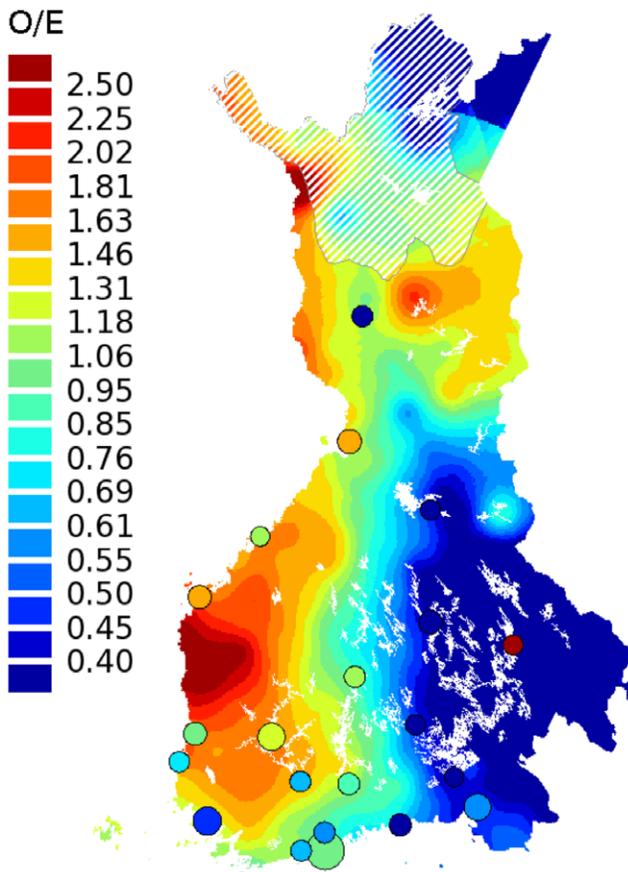


Figure 3. The observed/expected (O/E) ratios of Kaposi sarcoma (KS) incidence according to municipality of birth. Persons born in rural Western Finland have much higher incidence of KS than persons born in Eastern Finland. Twenty-one largest cities are indicated by circles. The white lining is superimposed on the colors of the sparsely populated areas in Northern Finland (less than 1 inhabitant per km²). The borders of Finland are based on those before year 1940. doi:10.1371/journal.pone.0055209.g003

tion of KS patients is seen in Northeastern Finland, while Eastern Finland displays much less KS cases than expected by chance. Interestingly, the family of five patients (Fig. 2) originates from Eastern Finland, whereas most of the other ascertained familial cases originate from Western Finland (Fig. 1).

Human herpesvirus 8, also known as Kaposi sarcoma-associated herpesvirus, is a well-known etiologic factor for KS. To study whether the uneven clustering of KS cases resulted from the uneven distribution of HHV8 infections in Finland, we studied the seroprevalence of HHV8 in 400 individuals from Eastern and Western Finland. Four out of 400 individuals (1%) were HHV8 positive. Two of them originated from Eastern Finland and two from Western Finland, suggesting similar seroprevalence of HHV8 in Eastern and Western Finland. However, the numbers are far too small for a proper conclusion on possible geographical variation of HHV8 infection in Finland.

Discussion

Here we have used an approach to systematically analyze familial aggregation of all tumor types in nation-wide registry-based data. Unlike most traditional epidemiological tools, our approach was able to derive evidence for familial occurrence of

both common and rare tumor types. Tumor types showing the greatest potential for familial occurrence were those that are mainly seen in rare cancer predisposition syndrome patients - a proof of concept that the method works well in estimating familial occurrence. Unexpectedly, one of the top scoring tumor types was Kaposi sarcoma with strong familial aggregation and uneven geographical distribution. We were able to identify seven families with two first-degree relatives with KS, and a family with five KS cases, which to our knowledge is the third largest reported KS family worldwide [16,17].

The described clustering method is designed to identify cancer cases with similar phenotype and common family name at birth, thus the method is able to detect also potential relatives outside the nuclear family. The method is also applicable to other registered diseases than cancer. Primary requirement is to obtain information about family names and municipalities of patients, while population-scaled data is needed for calculation of statistical significance. Populations with enough diversity in family names are suitable for similar clustering efforts and allow the formation of statistically significant clusters. Due to the fact that family name in Finland is usually inherited from father, the birth name-municipality based clustering method is restricted to finding family links only from the paternal side.

In this study, we were able to cover significantly more cases and more detailed tumor types compared to previous familial clustering efforts, such as those conducted in Icelandic [18] or Utah population databases [19,20]. One limitation of our effort is the lack of such extensive genealogy data for the studied individuals as in the studies in Iceland or Utah. However, the diversity of family names in Finland (~100,000) and relatively little migration before World War II - the time period when most of the cancer cases in FCR were born - warrants our search for familial occurrence with MN-clustering. In our example case of KS, we were able to ensure familial relations through NPR and PRs for 70% of the cases in MN-clusters, and 88% of these clusters were of high statistical significance with the lower confidence limit greater than ten. Thus, calculating the statistical significance of clustering, and choosing clusters with the high O/E and CI-low values obviously increase the probability of identifying true relatives.

We concentrated on the most significant MN-clusters to estimate the familial occurrence of the tumor types using cluster score. Top two tumor types indicated known rare cancer syndromes, namely Von Hippel-Lindau syndrome and Multiple Endocrine Neoplasia type 2, suggesting that our approach is feasible in identifying families with high-penetrance cancer predisposition. One of the most interesting phenotypes with a high cluster score was KS, which showed high tendency to cluster with a frequency exceeding, for example, that seen in Wilm's tumor. Another phenotype among the top tumor types without known genetic etiology was neuroendocrine carcinoma of small intestine, for which high familial risk has been described earlier [3,21].

Twin studies and systematic registry-based studies of all cancers have been restricted to demonstrate the hereditary effect concerning the most common cancer sites. In our study, the strongest evidence for familial occurrence among the most common tumor types was shown for papillary thyroid adenocarcinoma, chronic lymphatic leukemia, and squamous cell carcinoma of the lip, each with over 7000 registered patients in FCR. The very same cancer sites, although without detailed morphology definitions, showed significant relatedness also in distant familial relationships (no closer than first cousins) in Utah [20].

KS is a prevalent tumor type especially among HIV-infected patients and in sub-Saharan Africa [22]. HHV8 is a well-established etiologic factor for KS, and it is well acknowledged that

patients suffering from immunosuppressive conditions have higher risk for KS [23,24]. A few reports exist on familial occurrence of KS [16,17,25–28]. Accumulating evidence suggests genetic background as a predisposing factor to KS. A number of studies have associated certain human leukocyte alleles, and genetic variants in other host immunity related genes, with predisposition to KS [29–35]. However, no conclusive mutations predisposing to familial KS have thus far been reported. KS is a vascular tumor with an average incidence of 0.1–0.2 per 100,000 person-years in Finland in 1963–2010, adjusted for age to the World Standard Population. The incidence of KS according to municipality of birth varied strongly in Finland. Most KS patients were born near the west coast, leaving Eastern Finland with significantly fewer KS cases (Fig. 3). It is well reported, in a number of seroepidemiological studies, that HHV8 infection rates differ geographically [36]. Whether the west-east difference in the incidence is due to prevalence of the virus or local founder effect remains to be studied more carefully. Our preliminary serological analyses showed that the HHV8 infection is rare in Finland, and does not seem to be differentially distributed. HHV8 seroprevalence has not been studied in Finland earlier, whereas Finland's neighboring countries, Sweden in the west and Russia in the east, show seroprevalences of 10–20% and <10%, respectively [22]. The risk of HHV8 infection is higher among family members of classic KS patients, and person-to-person transmission has been confirmed by the observation that in almost all cases the same viral genotype was shared within the family [37,38]. While intrafamilial transmission of HHV8 probably plays a role in the familial occurrence of KS in our data, it cannot explain in itself the number of KS patients as high as five within a single family without additional susceptibility to develop tumors.

It is likely that some high-penetrance tumor predisposing genes – particularly those of recessive inheritance and those that do not display easily recognizable syndromic features – remain unidentified. Poorly characterized combinations of multiple germline variants with moderate penetrance are likely to explain some proportion of the hereditary cancer burden. While such genetic susceptibility factors have previously been difficult to study, the molecular tools available for hereditary cancer gene identification have improved dramatically in recent years. Next-generation

sequencing has highlighted the utility of very small numbers of carefully chosen affected individuals in genetic analyses [35,39,40]. Our approach is useful in deriving the most valuable specimens from the whole population for analyses of new cancer susceptibility conditions, including archival tissue blocks that can be utilized for generation of high-throughput DNA data for predisposition variant detection.

Supporting Information

Table S1 ICD-O-3 based morphology code groups used in the birth name-municipality based clustering.
(DOCX)

Table S2 ICD-O-3 based topography code groups used in the birth name-municipality based clustering.
(DOCX)

Table S3 Topography and morphology combinations in the birth name-municipality based clustering, distribution of patients, and cluster scores.
(DOCX)

Acknowledgments

The authors would like to acknowledge Kari Pasanen (the University of Eastern Finland) for developing the map illustration method, Risto Sankila (the Finnish Cancer Registry) and Ari Ristimäki (University of Helsinki) for expertise in pathology, Hilikka Laasanen (the Finnish Cancer Registry) for complex data processing assistance, Sini Nieminen, Sirpa Soisalo, Henna Degerlund and Minna Merikivi for genealogical research, Harri Rissanen and Laura Lund (National Institute of Health and Welfare) for help in obtaining the serum samples, and Maija Lappalainen (HUSLAB) for expertise in virology. We also thank Tiia Pelkonen for proofreading and editing the final text.

Author Contributions

Conceived and designed the experiments: EK M. Aavikko PV RL LAA E. Pukkala. Performed the experiments: EK TP E. Pukkala. Analyzed the data: EK M. Aavikko PV YL SS OK E. Pitkanen M. Artama RL LAA E. Pukkala. Contributed reagents/materials/analysis tools: PK ML. Wrote the paper: EK M. Aavikko PV TP SS OK E. Pitkanen M. Artama LAA E. Pukkala.

References

- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, et al. (2000) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343: 78–85.
- Blumenthal DT, Cannon-Albright LA (2008) Familiality in brain tumors. *Neurology* 71: 1015–1020.
- Hiripi E, Bermejo JL, Sundquist J, Hemminki K (2009) Familial gastrointestinal carcinoid tumours and associated cancers. *Ann Oncol* 20: 950–954.
- Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O (2009) Highly increased familial risks for specific lymphoma subtypes. *Br J Haematol* 146: 91–94.
- Hemminki K, Tretli S, Olsen JH, Tryggvadottir L, Pukkala E, et al. (2010) Familial risks in nervous system tumours: Joint Nordic study. *Br J Cancer* 102: 1786–1790.
- Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R (2007) Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res* 13: 356–361.
- Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, et al. (2008) The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet* 83: 787–794.
- de la Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations: The example of Finland revisited. *Proc Natl Acad Sci U S A* 95: 12416–12423.
- Vierimaa O, Georgitsi M, Lehtonen R, Vahteristo P, Kokko A, et al. (2006) Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science* 312: 1228–1230.
- Georgitsi M, Raitila A, Karhu A, Tuppurainen K, Makinen MJ, et al. (2007) Molecular diagnosis of pituitary adenoma predisposition caused by aryl hydrocarbon receptor-interacting protein gene mutations. *Proc Natl Acad Sci U S A* 104: 4101–4105.
- Teppo L, Pukkala E, Lehtonen M (1994) Data quality and quality control of a population-based cancer registry. experience in Finland. *Acta Oncol* 33: 365–369.
- Pukkala E (2011) Biobanks and registers in epidemiologic research on cancer. *Methods Mol Biol* 675: 127–164.
- Matikainen MP, Sankila R, Schleutker J, Kallioniemi OP, Pukkala E (2000) Nationwide cancer family ascertainment using Finnish cancer registry data on family names and places of birth for 35,761 prostate cancer patients. *Int J Cancer* 88: 307–312.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 289–300.
- Pukkala E, and Patama T (2010) Small-area based map animations of cancer incidence in Finland, 1953–2008. Available <http://astra.cancer.fi/cancermaps/suomi5308/>. Accessed 16 July 2012.
- Cottoni E, Masia IM, Masala MV, Mulargia M, Contu L (1996) Familial Kaposi's sarcoma: Case reports and review of the literature. *Acta Derm Venereol* 76: 59–61.
- DiGiovanna JJ, Safai B (1981) Kaposi's sarcoma. Retrospective study of 90 cases with particular emphasis on the familial occurrence, ethnic background and prevalence of other diseases. *Am J Med* 71: 779–783.
- Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, et al. (2004) Cancer as a complex phenotype: Pattern of cancer distribution within and beyond the nuclear family. *PLoS Med* 1: e65.

19. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86: 1600–1608.
20. Albright F, Teerlink C, Werner TL, Cannon-Albright LA (2012) Significant evidence for a heritable contribution to cancer predisposition: A review of cancer familiarity by site. *BMC Cancer* 12: 138.
21. Hemminki K, Li X (2001) Familial carcinoid tumors and subsequent cancers: A nation-wide epidemiologic study from Sweden. *Int J Cancer* 94: 444–448.
22. Mesri EA, Cesarman E, Boshoff C (2010) Kaposi's sarcoma and its associated herpesvirus. *Nat Rev Cancer* 10: 707–719.
23. Safai B, Johnson KG, Myskowski PL, Koziner B, Yang SY, et al. (1985) The natural history of Kaposi's sarcoma in the acquired immunodeficiency syndrome. *Ann Intern Med* 103: 744–750.
24. Farge D (1993) Kaposi's sarcoma in organ transplant recipients. The Collaborative Transplantation Research Group of Ile de France. *Eur J Med* 2: 339–343.
25. Hale LR, Kelly JW (1998) Familial Kaposi's sarcoma and Paget's disease of bone. *Aust J Derm* 39: 241–245.
26. Fata F, Mirza A, Bernath A, Bernath A (2002) Skin lesions in melanoma and Kaposi's sarcoma. Case 3. Familial classic Mediterranean Kaposi's sarcoma. *J Clin Oncol* 20: 1415–1418.
27. Guttman-Yassky E, Cohen A, Kra-Oz Z, Friedman-Birnbaum R, Sprecher E, et al. (2004) Familial clustering of classic Kaposi sarcoma. *J Infect Dis* 189: 2023–2026.
28. Weismann-Brenner A, Friedman-Birnbaum R, Brenner B (2004) Familial Kaposi's sarcoma: a cluster of five Israeli cases. *Clin Oncol* 16: 125–128.
29. Kaloterakis A, Papasteriades C, Filiotou A, Economidou J, Hadjiyannis S, et al. (1995) HLA in familial and nonfamilial Mediterranean Kaposi's sarcoma in Greece. *Tissue Antigens*, 45(2): 117–119.
30. Camcioglu Y, Picard C, Lacoste V, Dupuis S, Akcakaya N, et al. (2004) HHV-8-associated Kaposi sarcoma in a child with IFN γ R1 deficiency. *J Pediatr* 144: 519–523.
31. Gazouli M, Zavos G, Papaconstantinou I, Lukas JC, Zografidis A, et al. (2004) The interleukin-6-174 promoter polymorphism is associated with a risk of development of Kaposi's sarcoma in renal transplant recipients. *Anticancer Res* 24: 1311–1314.
32. Dorak MT, Yee LJ, Tang J, Shao W, Lobashevsky ES, et al. (2005) HLA-B, -DRB1/3/4/5, and -DQB1 gene polymorphisms in human immunodeficiency virus-related Kaposi's sarcoma. *J Med Virol* 76: 302–310.
33. Guerini FR, Agliardi C, Mancuso R, Brambilla L, Biffi R, et al. (2006) Association of HLA-DRB1 And DQB1 with classic Kaposi' sarcoma in mainland Italy. *Cancer Genomics and Proteomics* 3: 191–194.
34. Brown EE, Fallin D, Ruczinski I, Hutchinson A, Staats B, et al. (2006) Associations of classic Kaposi sarcoma with common variants in genes that modulate host immunity. *Cancer Epidemiol. Biomarkers Prev* 15: 926–934.
35. Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, et al. (2010) Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J Exp Med* 207: 2307–2312.
36. Dukers NH, Rezza G (2003) Human herpesvirus 8 epidemiology: what we do and do not know. *AIDS* 17: 1717–1730.
37. Guttman-Yassky E, Kra-Oz Z, Dubnov J, Friedman-Birnbaum R, Segal I, et al. (2005) Infection with Kaposi's sarcoma-associated herpesvirus among families of patients with classic Kaposi's sarcoma. *Arch Dermatol* 141: 1429–1434.
38. Mancuso R, Brambilla L, Agostini S, Biffi R, Hernis A, et al. (2011) Intrafamilial transmission of Kaposi's sarcoma-associated herpesvirus and seronegative infection in family members of classic Kaposi's sarcoma patients. *J Gen Virol* 92: 744–751.
39. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106: 19096–19101.
40. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al. (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 42: 30–35.