

1 **Cohort-based association study of germline genetic variants with acute and chronic**
2 **health complications of childhood cancer and its treatment: Genetic risks for childhood**
3 **cancer complications Switzerland (GECCOS) study protocol**

4 Nicolas Waespe, MD,^{1,2,3} Sven Strebel, MSc Pharm,^{1,2,4} Tiago Nava, MD,^{2,5} Chakradhara Rao
5 S. Uppugunduri, PhD Pharm,² Denis Marino MSc,² Veneranda Mattiello, MD,^{2,5} Maria Otth,
6 MD,^{1,3,6} Fabienne Gumy-Pause, MD,^{2,5} André O. von Bueren, MD, PhD,^{2,5} Frederic
7 Baleyrier, MD, PhD,^{2,5} Luzius Mader, PhD,¹ Adrian Spoerri, PhD, MPH,⁸ Claudia E. Kuehni,
8 MD, MSc^{1,7} Marc Ansari, MD.^{2,5}

9

10 ¹ Childhood Cancer Research Group, Institute of Social and Preventive Medicine, University
11 of Bern, Bern, Switzerland;

12 ² Research Platform for Pediatric Oncology and Hematology (CANSEARCH Research
13 Platform), Department of Pediatrics, Gynecology, and Obstetrics, University of Geneva,
14 Geneva, Switzerland;

15 ³ Graduate School for Cellular and Biomedical Sciences (GCB), University of Bern, Bern,
16 Switzerland;

17 ⁴ Graduate School for Health Sciences (GHS), University of Bern, Bern, Switzerland;

18 ⁵ Division of Pediatric Oncology and Hematology, Department of Women, Children, and
19 Adolescents, University Hospital of Geneva, Geneva, Switzerland;

20 ⁶ Division of Oncology-Hematology, Department of Pediatrics, Kantonsspital Aarau AG,
21 Aarau, Switzerland;

22 ⁷ Division of Pediatric Oncology and Hematology, Department of Pediatrics, University
23 Hospital of Bern, Bern, Switzerland;

24 ⁸ SwissRDL - Medical Registries and Data Linkage, Institute of Social and Preventive
25 Medicine, University of Bern, Bern, Switzerland.

26

27 **Corresponding author:** Prof. Dr. med. Marc Ansari, MD; Division of Pediatric Oncology
28 and Hematology, Department of Women, Children, and Adolescents, University Hospital of
29 Geneva, Rue Willy-Donzé 6; Bureau 5-507; 1211 Genève, Switzerland;

30 T: +41 22 372 47 31; F : +41 22 382 31 00

31 E-mail : marc.ansari@hcuge.ch

32

33

34 **Article metrics:**

35 Word count Abstract: **335**

36 Word count Main Text: **3510**

37 Number of Tables and Figures: total **5**

38 Number of Supporting Information files: **3**

39 References: **62**

40

41 **Short running title:** Genetic risks for childhood cancer complications Switzerland
42 (GECCOS)

43

44 **Background**

45 Childhood cancer and its treatment may lead to many acute and chronic health complications.
46 Related impairment in quality of life, excess in deaths, and accumulated health care costs are
47 relevant. There is a wide inter-individual variability in the type and severity of health
48 complications. Genetic variations are suggested to contribute to individual susceptibility. So
49 far, only few genetic variants have been used to risk-stratify treatment and follow-up care.
50 This study platform aims to identify germline genetic variants associated with acute and late
51 complications of childhood cancer.

52

53 **Methods**

54 The Genetic Risks for Childhood Cancer Complications Switzerland (GECCOS) study is a
55 nationwide cohort study. It includes patients and survivors who were diagnosed with
56 childhood cancers or Langerhans cell histiocytosis before age 21 years, were registered in the
57 Swiss Childhood Cancer Registry (SCCR) since 1976 and have consented to the Pediatric
58 Biobank for Research in Hematology and Oncology (BaHOP), Geneva, host of the Germline
59 DNA Biobank Switzerland for Childhood Cancer and Blood Disorders (BISKIDS). BISKIDS
60 is a national biobank for the collection of germline DNA in childhood cancer patients and
61 survivors.

62 GECCOS uses demographic and clinical data from the SCCR and the associated Swiss
63 Childhood Cancer Survivor Study (SCCSS), which contains health-related data of survivors.
64 Phenotypic data consist of objective measurements, health conditions diagnosed by
65 physicians, second primary neoplasms, self-reported and health-related information from
66 participants. Germline genetic samples and sequencing data have been collected in BISKIDS.
67 We will perform gene panel sequencing, whole-exome sequencing, or whole-genome

68 sequencing depending on the research questions. We will perform association analyses to
69 identify genetic variants associated with specified health conditions. We will use clustering
70 and machine-learning techniques and assess multiple health conditions in different models.

71

72 **Discussion**

73 GECCOS will serve as an overarching platform to enable genotype-phenotype association
74 analyses on complications associated with childhood cancer and its treatments. Knowledge of
75 germline genetic variants associated with childhood cancer-associated health conditions will
76 help to further individualize cancer treatment and follow-up care, potentially resulting in
77 improved efficacy and reduced side effects, for personalized cancer care.

78

79

80 **Trial registration**

81 Clinicaltrials.gov: NCT04702321

82

83

84 **Key words (max 3-10):** childhood cancer survivors, late effects, therapy complications,
85 second neoplasm, registry, biobank, genetic variation, genetic polymorphism,
86 pharmacogenetics, genetic association studies.

87

88

89 **MAIN MANUSCRIPT**

90 **1. Background**

91 Childhood cancers have become curable in $\geq 85\%$ of patients in developed countries.[1]
92 Current treatment protocols are multimodal with varying combinations of surgery,
93 chemotherapy, radiation, hematopoietic stem cell transplantation (HSCT), and
94 immunotherapies. The price of increased survival is a wide range of acute and chronic health
95 conditions. Cancer treatments are associated with acute complications such as transient
96 nausea and vomiting, mucositis, and fatigue but also pneumonitis, cardiomyopathy,
97 encephalitis, and life-threatening infections.[2] While many of these conditions are potentially
98 reversible, some are not and become chronic or develop over time like cardiac, pulmonary,
99 auditory, endocrine, reproductive, and neurocognitive health conditions, and second primary
100 neoplasms (SPNs).[3] A recent publication found a cumulative incidence of severe chronic
101 health conditions of 96% in childhood cancer survivors aged 50 years. The number of severe
102 chronic health conditions was 2 times higher in survivors compared to matched community
103 controls.[4] Mortality is significantly increased in survivors compared to the general
104 population.[5] Recurrence and SPNs are the leading causes of death in the first 2 decades after
105 cancer diagnosis followed by diseases of the cardiovascular and respiratory systems
106 thereafter.[5, 6] Because of their young age at diagnosis, survivors have decades of life time
107 ahead. The burden of chronic conditions, related impairment in quality of life, excess in
108 deaths, and accumulated health care costs is therefore of great relevance for them.[7]

109

110 Only few genetic variants modifying the risk of acute and chronic toxicities in children with
111 cancer have so far led to personalized treatment protocols or follow-up care.[8, 9] An example
112 is 6-mercaptopurine, where dosing is routinely adapted in contemporary treatment protocols
113 for patients with thiopurine methyltransferase (*TPMT*) variants which increase the risk of

114 acute hematological toxicity in the treatment of childhood acute lymphoblastic leukemia
115 (ALL).[10] Other genetic markers have been associated with various outcomes but so far not
116 implemented in treatment protocols, such as nudix hydrolase 15 (*NUDT15*) with
117 hematological toxicity after purine analogue treatment,[11, 12] dihydrofolate reductase
118 (*DHFR*) with overall survival in some subtypes of ALL,[13, 14] and glutathione S transferase
119 genes with various outcomes after hematopoietic stem cell transplantation.[15] Also, the
120 effect of genetic variation on late toxicities, chronic health conditions arising after the end of
121 childhood cancer treatment, has been investigated.[16, 17] Different clinical outcomes such as
122 hearing loss,[18–21] cardiomyopathy,[22, 23] metabolic syndrome,[24, 25] and gonadal
123 impairment have been studied.[26, 27] Exome-wide (EWAS) and genome-wide association
124 studies (GWAS) have led to the identification of genetic variants associated with drug
125 toxicities (e.g. osteonecrosis in children with ALL [*BMP7* and *PROXI-ASI*], asparaginase
126 hypersensitivity [*GRIAI*], or vincristine-associated peripheral neuropathy in children with
127 ALL [*CEP72*]),[28–31] metabolic syndrome,[24] sinusoidal obstruction syndrome after
128 HSCT,[32] and hearing loss.[18] Acylphosphatase 2 (*ACYP2*) was associated with hearing
129 loss after platinum treatment exposure in several independent datasets but failed to be
130 replicated in a recent large candidate gene analysis while a variant in solute carrier family 22,
131 member 2 (*SLC22A2*) was found to be associated with mild hearing loss.[20, 33] Genetic
132 modifiers have been implicated for the development of SPNs including breast cancer,[34]
133 CNS tumors,[35] and leukemia.[36] For other SPNs (such as thyroid cancer), no data is
134 available.[37] To address the contribution of genetic risk variants in the development of late
135 toxicities, large cancer survivor studies such as the Childhood Cancer Survivor Study (CCSS)
136 in the US are collecting DNA systematically to conduct genotype-phenotype analyses.[36]

137

138 Most studies on genetic risk variants used a candidate-gene approach, had a small sample size
139 of less than 200 participants, heterogenous cohorts with various treatment exposures, and
140 inconsistent outcome assessments.[17] Many health conditions have not been investigated,
141 like renal insufficiency, pulmonary, and ocular complications. It is likely that several distinct
142 pathways and their corresponding gene variants are involved in the development of complex
143 phenotypes like pulmonary dysfunction. Therefore, many candidate gene variants and
144 treatment exposures need to be considered. This is possible with hypothesis-free exome-wide
145 or genome-wide association studies (EWAS or GWAS). New analytical approaches which
146 combine clinical, pharmacological, and genetic data into integrative models have been
147 developed and are showing promising results.[38, 39] Network analyses, machine learning,
148 and clustering methods might help to understand the impact of genetic variations on complex
149 phenotypes with biological pathways and their corresponding genes. The GECCOS study will
150 enable large genotype-phenotype association studies in childhood cancer patients and
151 survivors in Switzerland and international collaborations. Selected organ dysfunctions and
152 second primary neoplasms will be studied in sub-projects, and *in silico* and *in vitro* methods
153 will be used to further explore mechanisms associated with genetic variation and outcomes.

154

155

156 2. Study objectives

157 The main objectives of the GECCOS study are

- 158 (i) to identify genetic variants associated with health conditions in childhood cancer
159 and its treatment using genotype-phenotype association methods;
- 160 (ii) to evaluate the biological function of genetic variants associated with health
161 conditions through *in silico* and *in vitro* methods; and

162 (iii) to identify genetic variants associated with multiple health conditions in childhood
163 cancer using models integrating multiple outcomes.

164 The secondary objectives are

165 (a) to create a common framework for sub-studies using genotype-phenotype
166 associations with germline genetic material and data of childhood cancer patients
167 and survivors;

168 (b) to feed germline genetic data generated in sub-studies into a biobank database for
169 future research and create a growing repository of genetic sequencing data; and

170 (c) to facilitate research using a common structure that can be used for collaborations.

171

172

173 **3. Methods**

174 **3.1. Study design**

175 The GECCOS study is a nationwide cohort study in collaboration of the Pediatric Biobank for
176 Research in Hematology and Oncology (BaHOP) Geneva, Switzerland, and the Institute for
177 Social and Preventive Medicine at the University of Bern. The Geneva Cantonal Commission
178 for Research Ethics has approved the GECCOS study (approval 2020-01723), and the BaHOP
179 biobank (approval PB_2017-00533).

180

181 **3.2. Data sources**

182 The GECCOS study uses genetic data and material from BaHOP. Clinical information is
183 collected from (i) the Swiss Childhood Cancer Registry (SCCR) and (ii) the Swiss Childhood
184 Cancer Survivor Study (SCSS; **Figure 1**). The nationwide germline DNA biobank for
185 childhood cancer and blood disorders (BISKIDS) was established in May 2019 within the

186 Biobank for Research in Hematology and Oncology (BaHOP) with support of all nine
187 pediatric oncology centers caring for childhood cancer patients in Switzerland. Return of
188 results and relevant incidental outcomes to the patient is defined in the BaHOP regulations
189 with oversight of a genetic advisory board. BISKIDS collects germline DNA samples,
190 extracts and stores genomic DNA and genetic data of childhood cancer patients as well as
191 survivors in Switzerland.

192

193 The SCCR collects information on childhood cancer patients diagnosed in Switzerland.[40]
194 Children and adolescents aged <21 years with a primary cancer diagnosis according to the
195 International Classification of Childhood Cancer, third edition (ICCC-3), and Langerhans cell
196 histiocytosis were registered since 1976 in the SCCR. The dataset is managed by the
197 Childhood Cancer Research Group at the Institute of Social and Preventive Medicine,
198 University of Bern. It has a completeness of coverage of childhood cancer patients in
199 Switzerland aged up to 15 years of >85% since 1985 and >95% since 1995.[41] As of
200 December 2019, 13,029 patients were registered in the SCCR, of which 9,306 (69%) were
201 still alive and potentially eligible for participation in GECCOS. Mandated by a national
202 cancer registration law in Switzerland enacted on January 1st, 2020, registration of new cancer
203 patients is performed by the federal government from that date onwards.[42] The Institute of
204 Social and Preventive Medicine of the University of Bern was commissioned to perform
205 patient registrations and data collection, and the Childhood Cancer Research Group will
206 continue research activities on these datasets. The GECCOS project will seek to include
207 patients diagnosed with neoplasms during childhood after January 1st, 2020, and continue
208 collaboration with the Childhood Cancer Research Group at the University of Bern. The
209 Swiss Childhood Cancer Survivor Study (SCCSS, [clinicaltrials.gov: NCT03297034](https://clinicaltrials.gov/ct2/show/study/NCT03297034)) is a
210 population-based, long-term cohort study of all childhood cancer patients who were registered

211 in the SCCR, resident in Switzerland at diagnosis, and survived ≥ 5 years after initial cancer
212 diagnosis.[43] The SCCSS collects questionnaire-based information from survivors, and data
213 on chronic health conditions after childhood cancer from survivorship clinics and hospital
214 records (e.g. audiograms, and lung function tests) which will be available for GECCOS.

215

216 3.3. Study population

217 Eligible in the participation in the GECCOS study are persons who:

- 218 1) are registered in the Swiss Childhood Cancer Registry; and
- 219 2) were diagnosed with a neoplasm according to the ICCC-3, or Langerhans cell
220 histiocytosis (LCH) before age 21 years; and
- 221 3) have consented themselves or through their legal representative to the BaHOP (host
222 biobank for BISKIDS).

223 Inclusion criteria for sub-projects focusing on specific health conditions will be defined
224 (**Supplementary Table 1 and 2** for established sub-projects). Within sub-projects, we will
225 sample patients and survivors according to risk exposure (for a cohort design), or according to
226 the outcome of interest (for a case-control or case-cohort design). We will identify eligible
227 participants from the SCCR and SCCSS.

228

229 3.4. Outcomes: health conditions and second primary neoplasms

230 We will assess health conditions in childhood cancer patients and survivors by collecting data
231 on organ function such as pulmonary functions tests for lung conditions, or audiograms for
232 hearing loss. For outcomes that cannot be adequately measured (e.g. tinnitus), we will use
233 information from self-assessment questionnaires. We will use the International Agency for

234 Research on Cancer (IARC) criteria for coding SPNs.[44] In brief, we will include neoplasms
235 according to the ICC-3 classification which originated in different tissues or had a different
236 morphology than the first primary neoplasm. We will not classify progression, transformation,
237 metastasis, and relapse of first primary neoplasm as SPN.

238

239 **3.5. Covariates**

240 For specific sub-studies and analyses, we will extract data on relevant covariates that might
241 influence the outcomes of interest from the SCCR and SCCSS. We will extract demographic,
242 socio-economic, first primary neoplasm, treatment, and follow-up information. For treatment
243 information, we will estimate cumulative doses of chemotherapies using individual treatment
244 protocols, or calculate effective treatment doses from medical records, if available. We will
245 estimate exposure to radiotherapy using radiotherapy field descriptions,[45–47] or calculate
246 organ-specific exposure from effective administered radiotherapy documentation (**Table 1**).

247

248 **3.6. Selection of participants**

249 We will identify participants eligible for specific sub-studies with defined in- and exclusion
250 criteria. We will use information in the SCCR and SCCSS and assess availability of
251 corresponding germline DNA samples or sequencing data from previously sequenced
252 participants in BISKIDS (**Figure 2**). If clinical data is available for a sufficient number of
253 participants but further genetic samples are needed, we will invite potential participants to
254 contribute germline DNA samples to BISKIDS for research. For collection of biological
255 material within BISKIDS, we will use two pathways: (i) invitations to participate are sent out
256 by the Childhood Cancer Research group at the University of Bern, consisting of germline
257 DNA collection kits (predominantly using saliva samples or buccal swabs) with information

258 on the biobanking project and associated research, and informed consents to the participant's
259 home; (ii) participants are invited by healthcare staff in hospitals caring for childhood cancer
260 patients and survivors. These potential participants and their legal representatives are
261 informed of the project and written consent and germline DNA are collected during a medical
262 visit already planned for their treatment or follow-up. All participants consent to have their
263 germline DNA stored in the BISKIDS section of the BaHOP biobank and their health-related
264 data to be used for genotype-phenotype association studies. All specific GECCOS sub-studies
265 will be reviewed and approved by a national scientific committee and submitted to the
266 responsible ethics committee as amendment to the main GECCOS protocol, insofar as the
267 applicable law requires authorization.

268

269 3.7. Data linkage

270 Information allowing to link genotype data from the BISKIDS collection with phenotype data
271 from the SCCR and SCCSS is securely stored in a separate trust center database managed by
272 a third party (SwissRDL – Swiss Medical Registries and Data Linkage, ISPM, University of
273 Bern). This procedure allows to separate identifying information from clinical and genetic
274 information. We use a web-based secured and personalized access point (currently RedCap
275 9.5, Vanderbilt University, Nashville, TN, USA). The key located at the trust center is used to
276 merge the clinical dataset with the germline genetic information without releasing identifying
277 information to the GECCOS research team or to the BaHOP biobank. The research dataset
278 will only contain a unique study-specific patient-identifier, without any identifying
279 information. Study-specific identifiers will be securely stored in the trust center database to
280 assure traceability of datasets (**Figure 3**).

281

282 3.8. Data and sample handling

283 Clinical data will be transferred in standard data formats. Analysis datasets will not include
284 identifying information. Management of germline DNA samples (e.g. saliva or buccal swabs),
285 DNA extraction, aliquoting, and storage procedures are clearly defined in the BaHOP biobank
286 regulation. For germline DNA sequencing in the GECCOS study, one of the DNA aliquots
287 will be sent to Campus Biotech, Geneva, a sequencing facility. For genotype-phenotype
288 analyses, we will collaborate with the Swiss Institute of Bioinformatics, Switzerland.
289 Sequencing data and relevant clinical outcome data will be shared in a secured and encrypted
290 way between the sequencing facility, the Swiss Institute of Bioinformatics for analysis, and
291 the research platform for pediatric oncology and hematology in Geneva. Data will be stored
292 and made available through a harmonized nationwide network to support computational
293 biomedical research and clinical bioinformatics.[48]

294

295 3.9. Sequencing analyses

296 We will use genetic information (**Table 2**) from different genomic sequencing methods: (I)
297 gene panel sequencing adapted to specific research questions (using TruSeq DNA PCR free
298 library preparation kit, Illumina, San Diego, USA); (II) whole-exome sequencing (using
299 Illumina HiSeq4000 or NovaSeq 6000 platform with a mean read depth of at least 70x); and
300 (III) whole-genome sequencing (with a mean read depth of at least 30x), depending on the
301 research question. We will use workflows for genotyping implemented in the Genome
302 Analysis Toolkit (GATK)[49] and adapted to the aim of the study for (1) sequence
303 generation; (2) sequence alignment; (3) variant calling; (4) variant filtering; and (5) variant
304 annotation.[50] We will also include data on read quality control. We will perform analyses
305 with any of the following: (i) a candidate gene approach with filtered variants in pre-selected
306 genes based on scientific hypotheses, (ii) hypothesis-free exome-wide, (iii) or genome-wide
307 association analysis, (iv) and multivariate approaches such as clustering methods or machine

308 learning to identify associations of genetic variants with outcomes of interest. We will
309 perform meta-analyses using combined cohorts of discovery and replication datasets and
310 previous studies reporting on the same genetic variants, where possible.

311

312 We will mainly use the software packages Stata (Stata Corporation, Austin, Texas), R (R
313 Foundation for Statistical Computing, Vienna, Austria), and PLINK[51] for analyses.
314 Depending on the sub-study and data availability, we will use different pipelines for quality
315 control, filtering, and annotation.[52] Statistical significance tests will be two-sided and
316 appropriate significance levels will be applied, adjusting for multiple testing where
317 appropriate taking into account clinical co-variates, where possible (Bonferroni method, False
318 Discovery Rate by Benjamini and Hochberg, or similar). Statistical uncertainty of estimates
319 will be expressed as 95% confidence intervals.

320

321 3.10. Power calculation

322 For each sub-study, we will calculate the power of the planned association analysis using
323 appropriate tools, e.g. the Genetic Association Study (GAS) Power Calculator
324 (https://csg.sph.umich.edu/abecasis/gas_power_calculator).[53] We will perform power
325 calculations that are appropriate for the intended analyses such as genome-wide, exome-wide,
326 or candidate-gene association studies. We will use for the different approaches the expected
327 number of variants after filtering and define the adjusted cut-off p-value appropriate for
328 multiple testing. We will estimate the sample size taking into account the outcome of interest
329 incidence, the expected relative risk for possible risk variants, and minor allele frequency
330 cutoff values using different models (dominant, additive, recessive, where appropriate).

331

332 3.11. ***In silico and in vitro analyses***

333 We will use computational (*in silico*) tools to estimate the effect of specific gene variants on
334 gene regulation, splicing and expression of proteins (e.g. PolyPhen, SIFT, Human Splicing
335 Finder, Matinspector).[54–57] We will choose *in silico* methods to identify genetic variants
336 associated with complex disease mechanisms. Examples of such models are clustering
337 methods including similarity network fusion[38] and PEGASUS[58]) and deep learning
338 methods to explore interactions between genes and outcomes of interest.[59] We will use
339 multiple outcomes in combination with multiple treatment exposures in suitable models to test
340 their association with genetic variant data and identify genetic variants associated with
341 multiple complications.[60, 61]

342 We will perform *in vitro* analyses using cell culture models relevant to the outcomes of
343 interest. We will treat cell lines with antineoplastic agents or irradiate them and then perform
344 transcriptomic analyses to identify genes that are differentially expressed after exposure to
345 specific treatment modalities. We will utilize adapted approaches depending on the outcomes
346 of interest to prioritize genes for further use in genotype-phenotype association studies.
347 Through differential analysis of change in gene expression after treatment exposure, we will
348 seek to identify candidate genes for further use in association studies. In hypothesis-free
349 analysis methods (WGS or WES) we will perform *in vitro* studies to clarify the biological
350 function of identified genes and genetic variants.[54, 55]

351

352 3.12. **Validation and replication**

353 We will seek to validate variants identified by next-generation sequencing that were
354 associated with the respective outcome of interest using a different method (e.g. Sanger
355 sequencing or real-time PCR). After successful validation, we will seek to proceed to

356 replication of identified variants in a targeted analysis of an independent cohort containing
357 similar outcome information as the one analyzed in the primary dataset.

358

359

360 4. Discussion

361 The GECCOS study will enable genotype-phenotype association studies focusing on various
362 health conditions in childhood cancer patients and survivors. The large inter-individual
363 variability in response to antineoplastic treatments and occurrence of early and late
364 complications is currently addressed mainly in a trial-and-error approach i.e. by delaying and
365 adjusting treatment after occurrence of complications. Follow-up care is stratified by
366 treatment exposure but not by genetic predictors.[62] The advantage of germline genetic risk
367 variants is that this knowledge can be assessed when the workup of the neoplasm is made and
368 then used early in the course of the treatment as they do not change over time. Such
369 knowledge would allow to personalize treatment and follow-up care for individual patients
370 before clinical signs of complications are present. Knowledge of genetic variants associated
371 with treatment response will help maximizing treatment effect while reducing the risk for
372 complications and finding the balance of treatment intensity in the light of increasing survival
373 in childhood cancers. Genetic predictors will improve individual counselling of patients and
374 their families and help developing individualized follow-up guidelines.

375

376 Assessing multiple outcomes taking into account multiple covariates including treatment
377 exposures will help identify particularly vulnerable patients. As identified in previous
378 research, many patients suffer from several complications after childhood cancer treatment.[4]

379 Finding genetic variants associated with increased risk for multiple health conditions will help

380 identifying gene variants that contribute to several organ system complications. This approach
381 might help identifying patients which could most benefit from treatment adaptation and
382 preventive measures to reduce complications.

383

384 GECCOS provides a legal and organizational platform on how to use sensitive genetic data
385 with clinical information in association studies. It establishes structures that can be used by
386 researchers for national and international collaborative studies. Germline genetic sequencing
387 data generated in the GECCOS sub-studies will be stored after completion in BISKIDS
388 hosted within the BaHOP biobank, Geneva University Hospital. Clinical data will remain in
389 the described databases and only be temporarily linked for research studies increasing data
390 safety. The populations of interest will overlap between sub-studies and sequenced datasets
391 generated from participants included in completed sub-studies will contribute with their
392 germline genetic data to subsequent studies. This growing resource will reduce costs for
393 future studies, where only DNA of a fraction of the participants will have to be sequenced.
394 We will favor whole genome and whole exome sequencing to create datasets that are not
395 restricted to a specific research question but can be used for further research. We will then be
396 able to address different questions with the same datasets. A further strength of GECCOS is
397 the availability of a large clinical dataset collected since 1976, curated and updated follow-up
398 information, and survival data regularly.

399

400 GECCOS will be limited due to the fact that Switzerland is a small country with a limited
401 number of possible participants, which we can recruit for our research despite the nationwide
402 and population-based sampling. This will be counteracted by international collaborations.
403 Another issue with research on childhood cancer complications is that these health conditions
404 are complex diseases with likely many mechanisms leading to a specific outcome. This makes

405 identification of specific gene variants difficult. Many findings from studies were not
406 replicated in independent datasets. Candidate-gene studies have particularly suffered from
407 this. We will also explore novel methods to cluster and associate gene variants with clinical
408 outcomes.

409

410 Our workflow combining a large dataset of clinical information with germline genetic data
411 will enable genetic research on patient populations within Switzerland and facilitate
412 collaborations with other research groups. As all childhood cancers are rare diseases by
413 definition of the world health organization (WHO) with less than 1 in 2,000 people being
414 affected, patient numbers are generally small. Research on rare childhood cancer subtypes or
415 specific rare outcomes is only possible through international collaborations. We will provide a
416 platform for these collaborations with the GECCOS study.

417

418 **Abbreviations table:**

419	ALL	Acute lymphoblastic leukemia
420	BaHOP	Biobank for Hematology and Oncology in Pediatrics
421	BISKIDS	Germline DNA Biobank Switzerland for Childhood Cancer and Blood
422		Disorders
423	CNS	Central nervous system
424	CPS	Cancer predisposition syndrome
425	DNA	Deoxyribonucleic acid
426	EWAS	Exome-wide association study
427	GECCOS	Genetic risks for childhood cancer complications Switzerland
428	GWAS	Genome-wide association study
429	HSCT	Hematopoietic stem cell transplantation
430	IARC	International Agency for Research on Cancer
431	ICCC-3	International Classification of Childhood Cancer, third edition
432	ISPM	Institute of Social and Preventive Medicine
433	LCH	Langerhans cell histiocytosis
434	SCCR	Swiss Childhood Cancer Registry
435	SCCSS	Swiss Childhood Cancer Survivors Study
436	SPN	Second primary neoplasm
437	WHO	World Health Organization

438 **Declarations**

439 **Ethics approval and consent to participate**

440 The Geneva Cantonal Commission for Research Ethics has approved the GECCOS study
441 (approval 2020-01723), and the BaHOP biobank (approval PB_2017-00533).

442 The **sponsor** of the GECCOS study is the University Hospital Geneva, Rue Willy-Donzé 6,
443 1211 Geneva 4, Switzerland; responsible person: Prof. Klara Posfay Barbe, e-mail:
444 klara.posfaybarbe@hcuge.ch

445 The current **version** of the GECCOS protocol when this manuscript was submitted: Version
446 1.0; May 28th, 2020. Changes in the protocol will be communicated to collaborators by the
447 principal investigator and the main study team, if relevant, and amendments made on
448 clinicaltrials.gov.

449

450 **Consent for publication**

451 Not applicable.

452

453 **Availability of data and materials**

454 The datasets that will be generated within GECCOS and/or analyzed during the study will not
455 be publicly available due to concerns for anonymity of patients with childhood cancer being a
456 rare disease. Datasets are available from the corresponding author on reasonable request for
457 specific projects.

458

459

460 **Competing interests**

461 The authors declare that they have no competing interests

462

463 **Funding**

464 This study is supported by the CANSEARCH Foundation for BISKIDS, the host biobank

465 BaHOP, the research study GECCOS, and salary support to Nicolas Waespe. Further funding

466 comes from the Swiss National Science Foundation (31BL30_185396), and Swiss Cancer

467 Research (KFS-4722-02-2019, KLS/KFS-4825-01-2019).

468 The funding bodies have no role in the design of the study and collection, analysis, and

469 interpretation of data and in writing this manuscript.

470

471 **Authors' contributions**

472 NW: Conceptualization, Design, Methodology, Writing - all stages, Visualization.

473 SSt: Methodology, Design, Writing - Reviewing and Editing.

474 TN, CRSU: Methodology, Writing - Reviewing and Editing.

475 DM, VM: Design, Writing - Reviewing and Editing.

476 MO, FGP, AOB, FB, LM: Writing - Reviewing and Editing.

477 CEK, AS: Design, Methodology, Writing - Reviewing and Editing.

478 MA: Supervision, Conceptualization, Design, Methodology, Writing - Reviewing and

479 Editing.

480

481 All authors have approved the submitted version and have agreed both to be personally
482 accountable for the author's own contributions and to ensure that questions related to the
483 accuracy or integrity of any part of the work, even ones in which the author was not
484 personally involved, are appropriately investigated, resolved, and the resolution documented
485 in the literature.

486

487 **Acknowledgements**

488 We thank all childhood cancer patients, survivors, and families for participating in our study.
489 We thank the study team of the Childhood Cancer Research Group, Institute of Social and
490 Preventive Medicine, University of Bern, and SCCSS (Christina Schindera, Tomas Slama),
491 the data managers of the Swiss pediatric oncology clinics (Claudia Althaus, Nadine
492 Assbichler, Pamela Balestra, Heike Baumeler, Nadine Beusch, Sarah Blanc, Pierluigi
493 Brazzola, Susann Drerup, Janine Garibay, Franziska Hochreutener, Monika Imbach,
494 Friedgard Julmy, Eléna Lemmel, Heike Markiewicz, Annette Reinberg, Renate Siegenthaler,
495 Astrid Schiltknecht, Beate Schwenke, and Verena Stahel) and the data managers and
496 administrative staff of the SCCR (Meltem Altun, Erika Brantschen, Katharina Flandera,
497 Elisabeth Kiraly, Verena Pfeiffer, Julia Ruppel, Ursina Roder, and Nadine Lötscher). We
498 thank the study team of the Research Platform for Pediatric Oncology and Hematology at the
499 Geneva medical school (Khalil Ben Hassine, Simona Jurkovic Mlakar, Fanny Muet, Mary
500 Khoshbeen-Boudal, Laurence Lesne, Vid Mlakar, Shannon Robin, Yoann Sarmiento) and the
501 Onco-Hematology Unit of the HUG (Fanette Bernard, Laurent Cimasoni, Violaine Guignon,
502 Nelly Hafner-Bénichou, Rodolfo Lo Piccolo).

503

504

505

506 **References**

- 507 1. Bhatia S, Armenian SH, Armstrong GT, van Dulmen-den Broeder E, Hawkins MM, Kremer LCM,
508 et al. Collaborative Research in Childhood Cancer Survivorship: The Current Landscape. *JCO*.
509 2015;33:3055–64.
- 510 2. Nurgali K, Jagoe RT, Abalo R. Editorial: Adverse Effects of Cancer Chemotherapy: Anything New
511 to Improve Tolerance and Reduce Sequelae? *Front Pharmacol*. 2018;9. doi:10.3389/fphar.2018.00245.
- 512 3. Hudson MM, Ness KK, Gurney JG, Mulrooney DA, Chemaitilly W, Krull KR, et al. Clinical
513 Ascertainment of Health Outcomes Among Adults Treated for Childhood Cancer. *JAMA*.
514 2013;309:2371–81.
- 515 4. Bhakta N, Liu Q, Ness KK, Baassiri M, Eissa H, Yeo F, et al. The cumulative burden of surviving
516 childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *The Lancet*.
517 2017;390:2569–82.
- 518 5. Schindler M, Spycher BD, Ammann RA, Ansari M, Michel G, Kuehni CE, et al. Cause-specific
519 long-term mortality in survivors of childhood cancer in Switzerland: A population-based study. *Int J*
520 *Cancer*. 2016;139:322–33.
- 521 6. Winther JF, Kenborg L, Byrne J, Hjorth L, Kaatsch P, Kremer LCM, et al. Childhood cancer
522 survivor cohorts in Europe. *Acta Oncologica*. 2015;54:655–68.
- 523 7. Guy GP, Berkowitz Z, Ekwueme DU, Rim SH, Yabroff KR. Annual Economic Burden of
524 Productivity Losses Among Adult Survivors of Childhood Cancers. *Pediatrics*. 2016;138 Supplement
525 1:S15–21.
- 526 8. Kovalchik SA, Ronckers CM, Veiga LHS, Sigurdson AJ, Inskip PD, de Vathaire F, et al. Absolute
527 risk prediction of second primary thyroid cancer among 5-year survivors of childhood cancer. *J Clin*
528 *Oncol*. 2013;31:119–27.
- 529 9. Chen Y, Chow EJ, Oeffinger KC, Border WL, Leisenring WM, Meacham LR, et al. Traditional
530 Cardiovascular Risk Factors and Individual Prediction of Cardiovascular Events in Childhood Cancer
531 Survivors. *J Natl Cancer Inst*. 2020;112:256–65.
- 532 10. McLeod HL, Coulthard S, Thomas AE, Pritchard SC, King DJ, Richards SM, et al. Analysis of
533 thiopurine methyltransferase variant alleles in childhood acute lymphoblastic leukaemia. *British*
534 *Journal of Haematology*. 1999;105:696–700.
- 535 11. Zhou H, Li L, Yang P, Yang L, Zheng JE, Zhou Y, et al. Optimal predictor for 6-mercaptopurine
536 intolerance in Chinese children with acute lymphoblastic leukemia: NUDT15, TPMT, or ITPA genetic
537 variants? *BMC Cancer*. 2018;18:516.
- 538 12. Schaeffeler E, Jaeger SU, Klumpp V, Yang JJ, Igel S, Hinze L, et al. Impact of NUDT15 genetics
539 on severe thiopurine-related hematotoxicity in patients with European ancestry. *Genetics in Medicine*.
540 2019;21:2145–50.
- 541 13. Dulucq S, St-Onge G, Gagné V, Ansari M, Sinnott D, Labuda D, et al. DNA variants in the
542 dihydrofolate reductase gene and outcome in childhood ALL. *Blood*. 2008;111:3692–700.
- 543 14. Ceppi F, Gagné V, Douyon L, Quintin CJ, Colombini A, Parasole R, et al. DNA variants in DHFR
544 gene and response to treatment in children with childhood B ALL: revisited in AIEOP-BFM protocol.
545 *Pharmacogenomics*. 2018;19:105–12.

- 546 15. Ansari M, Curtis PH-D, Uppugunduri CRS, Rezgui MA, Nava T, Mlakar V, et al. GSTA1
547 diplotypes affect busulfan clearance and toxicity in children undergoing allogeneic hematopoietic stem
548 cell transplantation: a multicenter study. *Oncotarget*. 2017;8:90852–67.
- 549 16. Bhatia S. Role of genetic susceptibility in development of treatment-related adverse outcomes in
550 cancer survivors. *Cancer Epidemiol Biomarkers Prev*. 2011;20:2048–67.
- 551 17. Clemens E, van der Kooi ALF, Broer L, van Dulmen-den Broeder E, Visscher H, Kremer L, et al.
552 The influence of genetic variation on late toxicities in childhood cancer survivors: A review. *Critical*
553 *Reviews in Oncology/Hematology*. 2018;126:154–67.
- 554 18. Xu H, Robinson GW, Huang J, Lim JY-S, Zhang H, Bass JK, et al. Common variants in ACYP2
555 influence susceptibility to cisplatin-induced hearing loss. *Nature Genetics*. 2015;47:263–6.
- 556 19. Thiesen S, Yin P, Jorgensen AL, Zhang JE, Manzo V, McEvoy L, et al. TPMT, COMT and
557 ACYP2 genetic variants in paediatric cancer patients with cisplatin-induced ototoxicity.
558 *Pharmacogenet Genomics*. 2017;27:213–22.
- 559 20. Vos HI, Guchelaar H-J, Gelderblom H, de Bont ESJM, Kremer LCM, Naber AM, et al.
560 Replication of a genetic variant in ACYP2 associated with cisplatin-induced hearing loss in patients
561 with osteosarcoma: *Pharmacogenetics and Genomics*. 2016;26:243–7.
- 562 21. Clemens E, Broer L, Langer T, Uitterlinden AG, de Vries ACH, van Grotel M, et al. Genetic
563 variation of cisplatin-induced ototoxicity in non-cranial-irradiated pediatric patients using a candidate
564 gene approach: The International PanCareLIFE Study. *Pharmacogenomics J*. 2020;20:294–305.
- 565 22. Singh P, Wang X, Hageman L, Chen Y, Magdy T, Landier W, et al. Association of GSTM1 null
566 variant with anthracycline-related cardiomyopathy after childhood cancer-A Children’s Oncology
567 Group ALTE03N1 report. *Cancer*. 2020;126:4051–8.
- 568 23. Wang X, Sun C-L, Quiñones-Lombraña A, Singh P, Landier W, Hageman L, et al. CELF4 Variant
569 and Anthracycline-Related Cardiomyopathy: A Children’s Oncology Group Genome-Wide
570 Association Study. *JCO*. 2016;34:863–70.
- 571 24. Wilson CL, Liu W, Yang JJ, Kang G, Ojha RP, Neale GA, et al. Genetic and clinical factors
572 associated with obesity among adult survivors of childhood cancer: A report from the St. Jude
573 Lifetime Cohort. *Cancer*. 2015;121:2262–70.
- 574 25. van Waas M, Neggers SJMM, Uitterlinden AG, Blijdorp K, van der Geest IMM, Pieters R, et al.
575 Treatment factors rather than genetic variation determine metabolic syndrome in childhood cancer
576 survivors. *Eur J Cancer*. 2013;49:668–75.
- 577 26. Romerius P, Giwercman A, Moëll C, Relander T, Cavallin-Ståhl E, Wiebe T, et al. Estrogen
578 receptor α single nucleotide polymorphism modifies the risk of azoospermia in childhood cancer
579 survivors. *Pharmacogenet Genomics*. 2011;21:263–9.
- 580 27. van Dorp W, van den Heuvel-Eibrink MM, Stolk L, Pieters R, Uitterlinden AG, Visser JA, et al.
581 Genetic variation may modify ovarian reserve in female childhood cancer survivors. *Hum Reprod*.
582 2013;28:1069–76.
- 583 28. Karol SE, Mattano LA, Yang W, Maloney KW, Smith C, Liu C, et al. Genetic risk factors for the
584 development of osteonecrosis in children under age 10 treated for acute lymphoblastic leukemia.
585 *Blood*. 2016;127:558–64.
- 586 29. Fernandez CA, Smith C, Yang W, Mullighan CG, Qu C, Larsen E, et al. Genome-wide analysis
587 links NFATC2 with asparaginase hypersensitivity. *Blood*. 2015;126:69–75.

- 588 30. Chen S-H, Pei D, Yang W, Cheng C, Jeha S, Cox NJ, et al. Genetic variations in GRIA1 on
589 chromosome 5q33 related to asparaginase hypersensitivity. *Clin Pharmacol Ther.* 2010;88:191–6.
- 590 31. Diouf B, Crews KR, Lew G, Pei D, Cheng C, Bao J, et al. Association of an Inherited Genetic
591 Variant With Vincristine-Related Peripheral Neuropathy in Children With Acute Lymphoblastic
592 Leukemia. *JAMA.* 2015;313:815–23.
- 593 32. Ansari M, Petrykey K, Rezgui MA, Del Vecchio V, Cortyl J, Ralph R-O, et al. Genetic
594 Susceptibility to Hepatic Sinusoidal Obstruction Syndrome in Pediatric Patients Undergoing
595 Hematopoietic Stem Cell Transplantation. *Biol Blood Marrow Transplant.* 2020;26:920–7.
- 596 33. Langer T, Clemens E, Broer L, Maier L, Uitterlinden AG, Vries ACH de, et al. Usefulness of
597 current candidate genetic markers to identify childhood cancer patients at risk for platinum-induced
598 ototoxicity: Results of the European PanCareLIFE cohort study. *European Journal of Cancer.*
599 2020;138:212–24.
- 600 34. Morton LM, Sampson JN, Armstrong GT, Chen T-H, Hudson MM, Karlins E, et al. Genome-
601 Wide Association Study to Identify Susceptibility Loci That Modify Radiation-Related Risk for Breast
602 Cancer After Childhood Cancer. *J Natl Cancer Inst.* 2017;109. doi:10.1093/jnci/djx058.
- 603 35. Hosking FJ, Feldman D, Bruchim R, Olver B, Lloyd A, Vijayakrishnan J, et al. Search for
604 inherited susceptibility to radiation-associated meningioma by genomewide SNP linkage
605 disequilibrium mapping. *Br J Cancer.* 2011;104:1049–54.
- 606 36. Knight JA, Skol AD, Shinde A, Hastings D, Walgren RA, Shao J, et al. Genome-wide association
607 study to identify novel loci associated with therapy-related myeloid leukemia susceptibility. *Blood.*
608 2009;113:5575–82.
- 609 37. Zidane M, Cazier JB, Chevillard S, Ory C, Schlumberger M, Dupuy C, et al. Genetic susceptibility
610 to radiation-related thyroid cancer: a systematic review. *Endocr Relat Cancer.* 2019.
- 611 38. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for
612 aggregating data types on a genomic scale. *Nat Meth.* 2014;11:333–7.
- 613 39. Mezlini AM, Goldenberg A. Incorporating networks in a probabilistic graphical model to find
614 drivers for complex human diseases. *PLoS Comput Biol.* 2017;13:e1005580.
- 615 40. Michel G, von der Weid NX, Zwahlen M, Adam M, Rebholz CE, Kuehni CE, et al. The Swiss
616 Childhood Cancer Registry: rationale, organisation and results for the years 2001-2005. *Swiss Med*
617 *Wkly.* 2007;137:502–9.
- 618 41. Schindler M, Mitter V, Bergstraesser E, Gumy-Pause F, Michel G, Kuehni CE, et al. Death
619 certificate notifications in the Swiss Childhood Cancer Registry: assessing completeness and
620 registration procedures. *Swiss Med Wkly.* 2015;145:w14225.
- 621 42. Le Conseil fédéral CS. RS 818.33 Loi fédérale du 18 mars 2016 sur l'enregistrement des maladies
622 oncologiques (LEMO). Loi fédérale sur l'enregistrement des maladies oncologiques. 2020.
623 <https://www.admin.ch/opc/fr/classified-compilation/20121618/index.html>. Accessed 20 Nov 2020.
- 624 43. Kuehni CE, Rueegg CS, Michel G, Rebholz CE, Strippoli M-PF, Niggli FK, et al. Cohort Profile:
625 The Swiss Childhood Cancer Survivor Study. *Int J Epidemiol.* 2012;41:1553–64.
- 626 44. Working Group Report. International rules for multiple primary cancers (ICD-0 third edition):
627 *European Journal of Cancer Prevention.* 2005;14:307–8.

- 628 45. Ng A, Brock KK, Sharpe MB, Moseley JL, Craig T, Hodgson DC. Individualized 3D
629 reconstruction of normal tissue dose for patients with long-term follow-up: a step toward
630 understanding dose risk for late toxicity. *Int J Radiat Oncol Biol Phys*. 2012;84:e557-563.
- 631 46. Maraldo MV, Jørgensen M, Brodin NP, Aznar MC, Vogelius IR, Petersen PM, et al. The impact of
632 involved node, involved field and mantle field radiotherapy on estimated radiation doses and risk of
633 late effects for pediatric patients with Hodgkin lymphoma. *Pediatr Blood Cancer*. 2014;61:717–22.
- 634 47. Gasic D, Rosenschöld PMA, Vogelius IR, Maraldo MV, Aznar MC, Nysom K, et al. Retrospective
635 estimation of heart and lung doses in pediatric patients treated with spinal irradiation. *Radiother
636 Oncol*. 2018;128:209–13.
- 637 48. Coman Schmid D, Cramer K, Oesterle S, Rinn B, Sengstag T, Stockinger H, et al. SPHN - The
638 BioMedIT Network: A Secure IT Platform for Research with Sensitive Human Data. *Stud Health
639 Technol Inform*. 2020;270:1170–4.
- 640 49. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome
641 Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.
642 *Genome Res*. 2010;20:1297–303.
- 643 50. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and
644 Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint
645 Recommendation of the Association for Molecular Pathology and the College of American
646 Pathologists. *The Journal of Molecular Diagnostics*. 2018;20:4–27.
- 647 51. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set
648 for whole-genome association and population-based linkage analyses. *Am J Hum Genet*.
649 2007;81:559–75.
- 650 52. Marees AT, Kluiver H de, Stringer S, Vorspan F, Curis E, Marie Claire C, et al. A tutorial on
651 conducting genome-wide association studies: Quality control and statistical analysis. *International
652 Journal of Methods in Psychiatric Research*. 2018;27:e1608.
- 653 53. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-
654 based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38:209–13.
- 655 54. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein
656 function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–81.
- 657 55. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations
658 using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
- 659 56. Desmet F-O, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing
660 Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37:e67.
- 661 57. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and
662 beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*. 2005;21:2933–
663 42.
- 664 58. Nakka P, Raphael BJ, Ramachandran S. Gene and Network Analysis of Common Variants Reveals
665 Novel Associations in Multiple Complex Diseases. *Genetics*. 2016;204:783–98.
- 666 59. Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn-Feuer A, Dinov ID, et al. Deep
667 learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics*.
668 2018;19:629–50.

669 60. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, et al. Multivariate and network
670 meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*.
671 2017;358. doi:10.1136/bmj.j3932.

672 61. Ristl R, Urach S, Rosenkranz G, Posch M. Methods for the analysis of multiple endpoints in small
673 populations: A review. *J Biopharm Stat*. 2019;29:1–29.

674 62. Landier W, Skinner R, Wallace WH, Hjorth L, Mulder RL, Wong FL, et al. Surveillance for Late
675 Effects in Childhood Cancer Survivors. *J Clin Oncol*. 2018;36:2216–22.

676

677

678

679 **Tables:**

680 **Table 1.** Summary information on main covariates and exposures of interest used in the

681 genotype-phenotype association analyses.

Covariate type	Covariates	Unit
Demographic information	Sex	Male/ female/ other
	Birthdate	Month/ year
	Country of origin	Country name
Socio-economic status	Highest education patient and parents	10-unit scale
	Income patient and/ or parents	Monthly net income (Swiss francs)
First primary neoplasm information	Age at diagnosis	Years
	Date of diagnosis	Month/ year
	Type of diagnosis	ICCC3 code; ICDO3 morphology, topography, behavior code
	Laterality	Left/ right/ bilateral/ medial/ not applicable
	Relapse date	Month/ year
	Relapse type	Local/ distant/ systemic/ other
	Relapse location	Organ and morphology
Treatment information	Treatment protocol	Name of protocol, arm, randomization group
Chemotherapy	Separately per antineoplastic agent: Cumulative dose estimated using treatment protocols or if available calculated using extracted data from medical records;	mg/m ² , or appropriate metric; cycles (n); dose per cycle (mg/m ²)
	Start date	Month/ year
Radiotherapy	Separately per radiation field: Cumulative dose estimated on an intention to treat basis using radiotherapy field descriptions, or calculated from radiotherapy protocols	Mean and maximal estimated dose per organ (Gy); fractions (n); dose per fraction (Gy)
	Radiation type	Photon, proton, brachytherapy, stereotactic radiation, other
	Radiation field	Description of radiation field (e.g. mantle field, whole lung irradiation, total body irradiation)

	Laterality	Left/ right/ bilateral/ medial/ not applicable
	Start date	Month/ year
	Concomitant chemotherapy	Antineoplastic agent and dose (mg/m ²)
Surgery	Location	Organ, site, and description of intervention
	Laterality	Left/ right/ bilateral/ medial/ not applicable
	Date	Month/ year
Hematopoietic stem cell transplantation	Type	Allogeneic, autologous, other
	Donor type	Matched sibling, matched other relative, haploidentical relative, matched unrelated, mismatched unrelated, other
	Donor graft source	Bone marrow, peripheral stem cells (apheresis), umbilical cord blood, other
	Matching degree	Number of HLA loci matched of total assessed HLA loci
	Conditioning regimen	Name and treatment details
	Total body irradiation	Yes (including dose in Gy), no
	Date of stem cell transfer	Month/ year
	Acute complications	Sinusoidal obstruction syndrome, infection, acute GvHD (with grading), others
	Chronic complications	Chronic GvHD (with grading)
Follow-up information	Last vital status	Alive, dead, unknown
	Date of last vital status	Month/ year
	Last follow-up information from clinical site	Month/ year

682

683 Legend: GvHD, graft versus host disease; Gy, Gray; HLA, human leukocyte antigen; ICC3,

684 International Classification of Childhood Cancer, third edition; ICDO3, International

685 Classification of Diseases for Oncology, third edition; mg/m², milligrams per square meter; n,

686 number.

687

688

689

690

691 **Table 2.** Summary information on the genetic information used in the genotype-phenotype
692 association analyses.

Information	Data	Format
Germline genetic sequencing data	Raw sequencing data	FASTQ or BAM files
	Identified genetic variants	VCF files
	Documentation of sequencing procedure and analyses performed to allow trackability of downstream analyses	CSV or TXT files to allow long-term readability
	Quality measures	Read depth, score to evaluate correct variant calling, etc.
Underlying genetic condition	Heritable underlying condition	Description of underlying condition
	Gene name of affected gene	HGVS name
	Identified causal variant	HGVS notation: genomic location, coding DNA, and expected effect on protein; rs number if available

693

694 Legend: BAM, compressed binary file used to represent aligned sequences; CSV, comma
695 separated values; DNA, deoxyribonucleic acid; FASTQ, text file containing the sequence data
696 from the clusters that pass filter on a flow cell; HGVS, Human Genome Variation Society; rs,
697 reference single nucleotide polymorphism identifier; TXT, standard unformatted text
698 document; VCF, variant call format.

699

700

701

702

703 **Figure titles and legends:**

704 **Figure 1.** Responsible teams in the GECCOS study for germline genetic association with
705 health conditions in childhood cancer patients and survivors.

706

707 Legend: BaHOP, Biobank for Hematology and Oncology in Pediatrics (BaHOP); BISKIDS,
708 Germline DNA Biobank for childhood cancer and blood disorders; DNA, deoxyribonucleic
709 acid; SCCR, Swiss Childhood Cancer Registry; SCCSS, Swiss Childhood Cancer Survivor
710 Study

711

712

713

714 **Figure 2.** Flow diagram of identification of eligible participants for a specific sub-project: (1)
715 The eligibility criteria, as defined by the researchers, will be transmitted to the Childhood
716 Cancer Research Group at the ISPM, University of Bern, and the BISKIDS biobank at the
717 Geneva University Hospital; (2) the number of eligible participants is compiled by secured
718 data exchange from the BISKIDS project and the Childhood Cancer Research Group; (3) The
719 number of eligible participants will be transmitted to the researcher to assess feasibility of a
720 sub-project.

721

722

723

724 **Figure 3.** Flow diagram of release of information from the different resources. (1)
725 Submission of sub-project to the responsible ethics committee, either as amendment to the
726 main protocol of GECCOS or as a separate project; (2) Submission of sub-project to the
727 scientific committee of the BaHOP, host biobank for the BISKIDS section at the Geneva
728 University Hospital; (3) Release of linking information from SwissRDL; (4) Transmission of
729 BISKIDS and SCCR identifiers for included patients with a newly generated study-specific
730 patient-identifier to BISKIDS and the ISPM respectively, to release variables used for the
731 study; (5) Transfer of data for included participants with a data transfer agreement to the
732 researcher with study-specific identifier; (6) After conclusion of the project: Storage of
733 acquired germline genetic data in the BISKIDS biobanking database at the Geneva University
734 Hospital for future research projects.
735

medRxiv preprint doi: <https://doi.org/10.1101/2021.01.30.21250679>; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

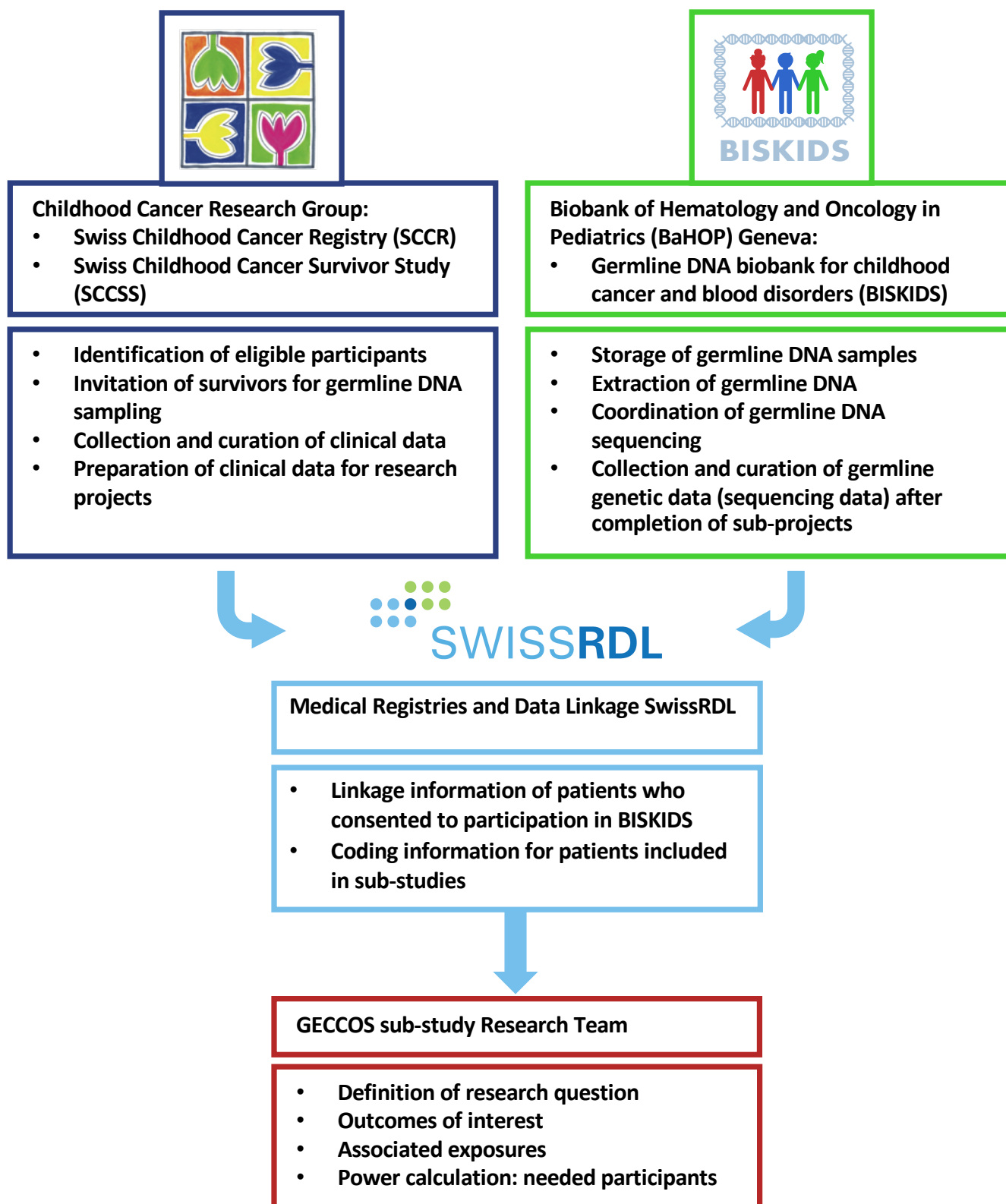


Figure 1

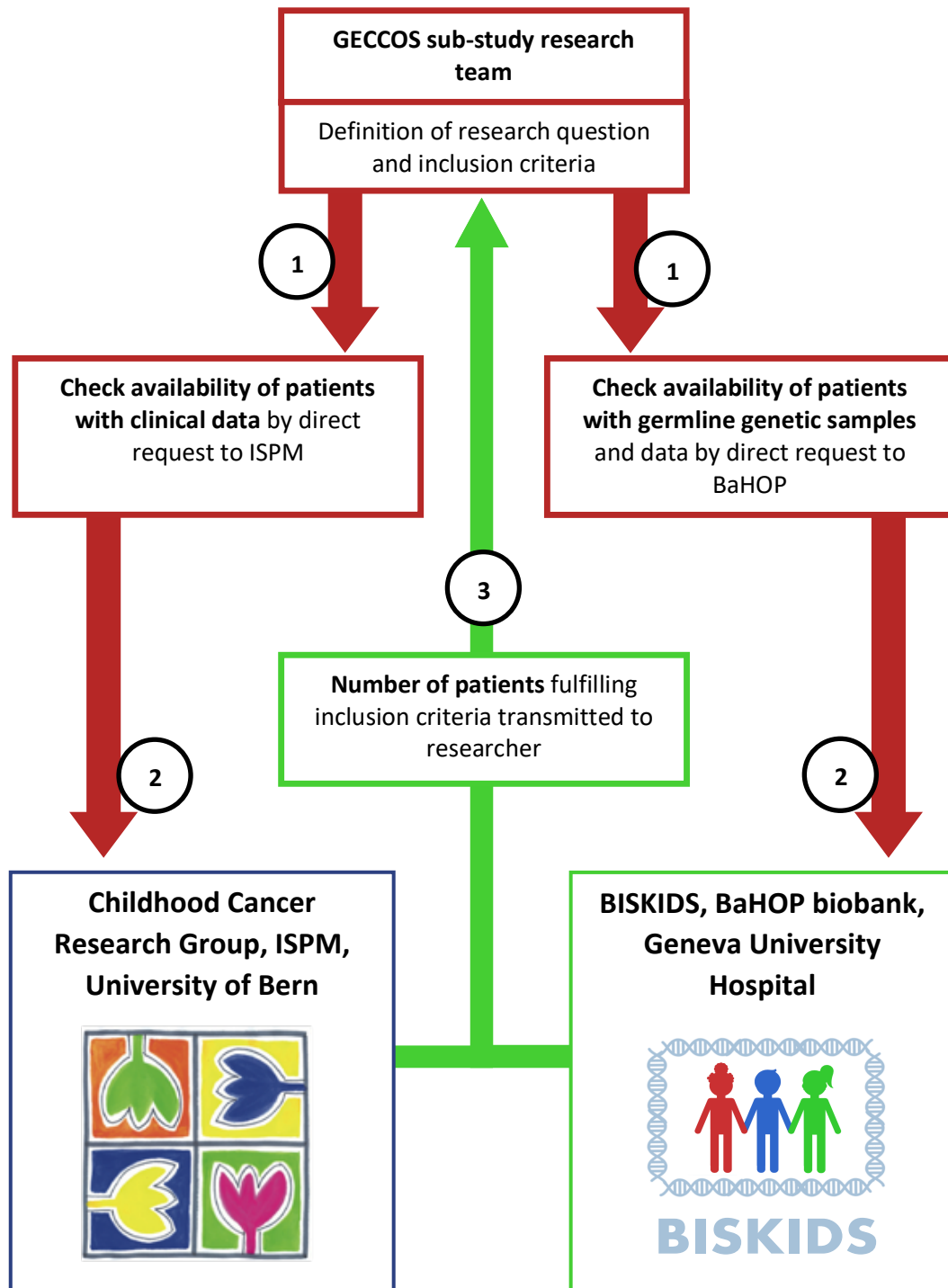


Figure 2

medRxiv preprint doi: <https://doi.org/10.1101/2021.01.30.21250679>; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

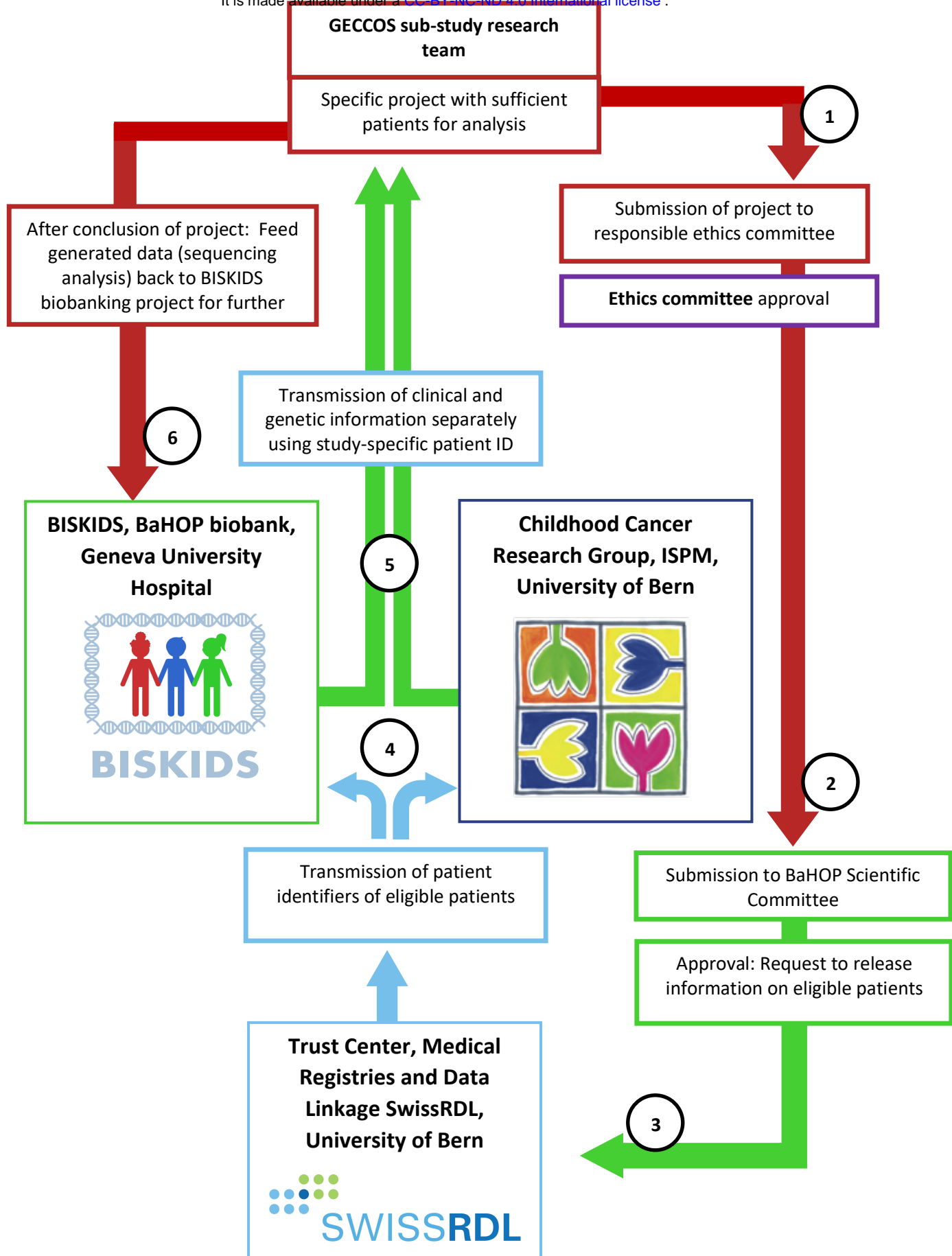


Figure 3