Protocol Synopsis (350 words maximum) :

PROJECT TITLE	**The molecular basis of developmental disorders: The case of uterovaginal aplasia**
BACKGROUND	The cause of congenital uterovaginal aplasia, or Mayer-Rokitansky-Küster-Hauser syndrome is unknown. The abnormalities could either be due to somatic genomic variants or germline mutations with reduced penetrance or sex-specific expression.
PRIMARY OBJECTIVE	To identify candidate genes for the MRKH syndrome through the investigation of genetic somatic differences between monozygotic twins discordant for this disease, and through the mutation spectrum comparison between uterine tissue remnants and matching blood DNA.
INCLUSION/ EXCLUSION CRITERIA	Patients with a diagnosis of MRKH syndrome properly characterized and evaluated by the Gynecology Department of the HUG that accept to donate blood and tissue from resected uterine buds, with a special interest in monozygotic twins discordant for this phenotype. First degree relatives of individuals with MRKH syndrome.
STUDY DESIGN	Whole genome high throughput sequencing (HTS) of monozygotic twins discordant for MRKH syndrome. Whole exome HTS of tissue/blood sets of MRKH patients. Low mosaicism screening of candidate genes/variants identified with whole genome or exome sequencing in a collection of blood and tissue samples of patients with MRKH syndrome with targeted high-depth HTS.
SAMPLE SIZE CALCULATION	The disorder is rare and thus we propose to study 5 monozygotic twin pairs and about 100 sporadic cases (and parents) and 25 tissue/blood sets. This sample size would be sufficient to test our hypotheses.
ENDPOINTS	We expect to identify strong candidate genes for the MRKH syndrome and provide a better understanding of the molecular mechanisms underlying this disease. This could only be achieved by the proposed collaboration between the clinical team at the HUG and the research group at CMU.
STATISTICAL ANALYSIS	Bioinformatic tools for the evaluation of variants have been developed by our group. The computational handling of the data will be performed using appropriate software for genomic analysis.
TIME SCHEDULE/STUDY DURATION	We propose a two year study. Patient assessment and sample collection will be performed during the first 14 months; sample preparation and sequencing throughout the first 16 months; data analysis and interpretation between months 6 and 24. The last six months will be devoted to integration of results and preparing the scientific reports.