WTSI Human Data Security Policy (October 2015)

I. **Scope** - This document outlines the policy for the processing of human genetics and/or genomics data (HGD) undertaken by WTSI researchers and/or using WTSI systems. The policy applies to all research programmes. Group leaders (PIs) responsible for projects involving HGD must ensure their teams are aware of this policy. Separate documents provide additional background[1] (including a risk analysis) and detailed implementation notes[2] and will be updated as risk profile and technical methods change over time.

II. **Assessment** - All projects involving HGD must be assigned to one of four data security levels (set out in section IV) prior to the start of data processing (and ideally prior to receipt or generation of data). The assignment should be carried out by the project PI and should be re-assessed at least once a year to ensure that it remains appropriate. If a project contains distinct datasets that would fall into multiple levels, there are two choices: the entire project can be assigned to the highest applicable level, or the HGD can be divided into multiple projects, each of which contains data only at the assigned level or below.

III. **Reporting breaches** – if HGD is processed in a way that contravenes with this policy, the matter must be reported to the Head of Legal. The Head of Legal will provide an annual report to HMDMC on reported breaches.

IV. **Data Security Levels** – The requirements of each of the four levels are described below. Requirements are **cumulative** (e.g. a project at level 3 must comply with the requirements of levels 1, 2, and 3).

Level 1: Open
Participants in studies producing entirely level 1 data will have given informed consent to make their data public without restriction (e.g. data from HapMap or 1000 Genomes Project). In addition, subsets of HGD may be handled at level 1 if all such subsets drawn from a particular study have been or could be disclosed publicly (e.g. in a talk or publication), even if the subsets are drawn from a study which in its complete form must be handled at a higher level (e.g. individual genotype data at a very small number of sites or genome-wide summary statistics drawn from a large population).
This policy does not impose restrictions on this type of data.

Level 2: Standard
We expect most projects will fall into this category. It consists of projects which include HGD that cannot be assigned to level 1 (e.g. because participants only consented to use of their data under certain conditions) and which do not include any Level 3 or 4 data. HGD at this level will typically be linked only to limited demographic/phenotypic information and be drawn from a relatively large population (e.g. data from WTCCC).
Register: Projects including HGD should be listed on a register held by each research programme. The register entry should include a brief description of the data set, its data security level, the name of the PIs, and any access agreements associated with the data. **Do not re-identify**: No attempt must ever be made to identify individuals to whom the data relates. **Intake**: HGD received from external sources should be checked to verify it complies with this policy. **Access Control Restrictions**: All systems (including computers, storage systems, portable devices, and the users and administrators who control that equipment) which are used to store, process, or transmit unencrypted HGD must limit access to authenticated individuals who are members of an access control group designated by a PI (and to WTSI systems administrators). **Encryption**: HGD stored on (or transmitted via) systems which do not implement the required

access control restrictions must be suitably encrypted. All associated decryption keys must be handled as if they were the original unencrypted HGD.
Direct Transfer: HGD may be directly transmitted between members of an access control group or, with the express permission of a PI, to any other individual (including those external to WTSI). Such transfers should be over a private/protected channel (e.g. direct USB cable, telephone), and not through any intermediary (e.g. email).
Disposition: Each copy of HGD should be deleted after it is no longer required (possibly never). Hardware that is removed from service that was previously used to store HGD should be wiped or destroyed before disposal.

Level 3: Strong

Stronger security is required when conditions of consent and/or data access agreements explicitly require it, when the data includes information that poses a greater risk of identification of the individuals involved, or when being identified as a member (or a close relative) of the population being studied would be particularly undesirable. In particular, this would include HGD that is linked to more extensive demographic/phenotypic information or which is drawn from populations for which the risks associated with re-identification are substantially higher (e.g. data from DDD).

Access Control Restrictions: Each member of the designated access control group(s) must be expressly authorised to access the HGD by a PI. Members must take precautions to ensure that any systems on which unencrypted HGD is stored remain under their (or system administrators') control.
Access Control Group Changes: Changes to access control group membership must be approved by a PI or their delegate, and all such changes should be recorded. PIs should review access control group membership at least once every 6 months and remove any individuals who no longer require access, and should ensure that individuals who are removed are informed of the need to destroy any copies of the data that they hold under their control.
Encryption: The encryption system should be as secure as practicably possible.

Level 4: Personal

Projects that include data that qualify as "personal data" under the Data Protection Act 1998 . This includes directly identifying data such as name and address (including email) but also covers data from which the individual could be identified if linked with other information that is reasonably available to the people accessing the data. If in doubt, consult the Legal department.

Access Control Restrictions: All systems used to store or process unencrypted HGD must either limit access to the entire system only to the members of the access control group for a single project or utilise a system-wide mandatory access control (MAC) policy which effectively isolates projects. The authentication system should use different credentials than those used for any other system, and should include a higher level of authentication such as by using two-factor authentication. Transmission of unencrypted HGD is not permitted between systems, and remote access to the system must be made only over encrypted links. Portable systems must not be used to store unencrypted HGD unless they are kept physically locked in a secure location at all times.
Legal: The Data Protection Act will apply and advice should be taken from Legal and IT Security as to how to conform to its requirements.

[1] "WTSI Human Genetics Data Security Policy: Background and Risk Analysis"
[2] "WTSI Human Genetics Data Security Policy: Implementation Notes"