

# Tabla de Contenidos

---

<b>Capítulo 1. Motivación .....</b>	<b>35</b>
1.1 Descripción del Problema.....	37
1.2 Objetivos de la tesis .....	38
1.3 Solución propuesta.....	39
1.4 Metodología de la Investigación .....	40
1.4.1 Framework Metodológico .....	40
1.4.2 Metodología aplicada a la tesis.....	42
1.5 Estructura de la Tesis Doctoral .....	45
<b>Capítulo 2. Dominio Genómico .....</b>	<b>48</b>
2.1 Genoma Humano.....	49
2.2 Secuenciación de Genomas .....	55
2.2.1 Pruebas Genéticas .....	56
2.2.2 ¿Qué es la secuenciación de exomas? .....	59
2.2.3 Tecnologías de Secuenciación.....	60
2.3 Background: Medicina de Precisión .....	62
2.4 Conclusiones .....	65
<b>Capítulo 3. Estado del Arte.....</b>	<b>67</b>
3.1 Modelado Conceptual en el Dominio Genómico.....	68
3.2 Bases de Datos Genómicas .....	70
3.2.1 1000 Genomas .....	72
3.2.2 ALFRED.....	73
3.2.3 BIC (Breast Cancer Information Core).....	75
3.2.4 BioQ .....	76
3.2.5 ClinVar .....	77
3.2.6 COSMIC.....	78
3.2.7 dbGAP.....	80

3.2.8	dbSNP .....	81
3.2.9	D-HaploDB (Definitive Haplotype Database) .....	82
3.2.10	DisGeNET .....	83
3.2.11	Ensembl.....	84
3.2.12	HapMap .....	85
3.2.13	HGMD.....	87
3.2.14	KEGG .....	88
3.2.15	LOVD.....	90
3.2.16	OMIM.....	92
3.2.17	REACTOME .....	93
3.2.18	SNPedia.....	94
3.2.19	UCSC.....	95
3.2.20	UMD (Universal Mutation Databases).....	96
3.2.21	UniProt (Universal Protein).....	97
3.2.22	YHRD.....	99
3.3	Comentarios adicionales .....	101
3.4	Conclusiones .....	104
<b>Capítulo 4. Evolución del Modelo Conceptual del Genoma Humano...107</b>		
4.1	Modelo Conceptual del Genoma Humano, versión 1 .....	109
4.1.1	Gene-Mutation View .....	110
4.1.2	Genome View .....	113
4.1.3	Transcription View.....	115
4.2	MCGH versión 1.1.....	117
4.2.1	Phenotype View .....	118
4.3	Desde v1 a v2: MCGH v2 .....	121
4.3.1	Eliminación banco de datos -genomas individuales- .	122
4.3.2	Los elementos cromosómicos como unidades básicas de modelado .....	123

4.3.3	Modelado de SNPs .....	125
4.3.4	Introducción de los conocimientos relacionados con: Pathways .....	126
4.4	Descripción de Clases: MCGH v2 .....	127
4.1.1	Vista Estructural .....	127
4.1.2	Vista de Transcripción.....	130
4.1.3	Vista de Variaciones .....	137
4.1.4	Vista de Rutas Metabólicas .....	144
4.1.5	Vista de Fuentes de Datos y Bibliografía .....	149
4.5	Conclusiones .....	153
<b>Capítulo 5. Estrategia de Integración de Haplotipos al MCGH .....</b>		<b>155</b>
5.1	Antecedentes: Comprendiendo el concepto de Haplotipo – caso práctico: Sensibilidad al Alcohol-.....	157
5.2	Trabajos Relacionados .....	160
5.3	Modelado Conceptual de Haplotipos.....	164
5.3.1	Validación del Modelo Conceptual .....	170
5.3.2	Desarrollo de una Base de Datos de Haplotipos.....	173
5.4	Evolución de la BD según el Modelo Conceptual.....	180
5.5	Conclusiones .....	183
<b>Capítulo 6. Implementación.....</b>		<b>186</b>
6.1	Metodología SILE.....	187
6.1.1	Ejemplos de búsquedas en repositorios genómicos ....	190
6.2	Base de Datos del Genoma Humano (HGDB).....	194
6.4.1	Selección de los repositorios de datos.....	196
6.4.2	Módulo de carga (genética) .....	198
6.3	Ficheros VCF .....	203
6.4	VarSearch (VS-prototipo) .....	205
6.4.1	Arquitectura de VarSearch .....	208

6.4.2	Guía de uso VS.....	210
6.4.3	Trabajos Relacionados .....	214
6.5	Caso de Estudio: Explotación del conocimiento genómico a través de VS.....	217
6.5.1	Explotación de tecnologías NGS.....	217
6.5.2	Optimización del Tiempo.....	220
6.6	GenesLove.Me .....	222
6.5.1	Arquitectura GenesLove.Me .....	224
6.7	Conclusiones .....	227
<b>Capítulo 7. Conclusiones.....</b>		<b>229</b>
7.1	Contribuciones principales .....	229
7.2	Impacto de la tesis .....	232
7.2.1	Publicaciones .....	232
7.2.2	Proyectos académicos.....	235
7.2.3	Participación en la comunidad de modelado.....	236
7.3	Trabajo futuro .....	238
<b>Referencias Bibliográficas.....</b>		<b>241</b>
<b>Anexos .....</b>		<b>257</b>
Anexo A. Diccionario de Datos .....		259
Anexo B. Glosario .....		267

# Índice de figuras

---

<b>Figura 1.</b> Framework para el “Design Science” aplicado al MCGH.....	41
<b>Figura 2.</b> Design Science como un ciclo regulativo .....	43
<b>Figura 3.</b> Ciclos regulativos de esta tesis doctoral.....	44
<b>Figura 4.</b> Elementos del cuerpo humano .....	50
<b>Figura 5.</b> Cronología del genoma humano (1866-2012).....	53
<b>Figura 6.</b> Noticia diario “El País” (27-junio-2000).....	54
<b>Figura 7.</b> Evolución plataformas de secuenciación de alto rendimiento ...	60
<b>Figura 8.</b> El uso de cambios genéticos .....	63
<b>Figura 9.</b> Proyecto 1000 Genomas (website).....	73
<b>Figura 10.</b> ALFRED (website).....	74
<b>Figura 11.</b> BIC (website).....	75
<b>Figura 12.</b> BioQ (website).....	76
<b>Figura 13.</b> ClinVar (website).....	78
<b>Figura 14.</b> COSMIC (website) .....	79
<b>Figura 15.</b> dbGAP (website) .....	80
<b>Figura 16.</b> dbSNP (website).....	81
<b>Figura 17.</b> DisGeNET (website) .....	83
<b>Figura 18.</b> Ensembl (website) .....	85
<b>Figura 19.</b> HapMap (website).....	86
<b>Figura 20.</b> Poblaciones tratadas Proyecto HapMap (3era. Fase).....	86
<b>Figura 21.</b> HGMD (website) .....	87
<b>Figura 22.</b> KEGG (website) .....	89
<b>Figura 23.</b> LOVD (website). .....	91
<b>Figura 24.</b> OMIM (website) .....	93
<b>Figura 25.</b> REACTOME (website).....	94
<b>Figura 26.</b> SNPedia (website) .....	95
<b>Figura 27.</b> UCSC (website).....	96
<b>Figura 28.</b> UMD (website) .....	97
<b>Figura 29.</b> UniProt (website).....	98
<b>Figura 30.</b> YHRD (website) .....	100
<b>Figura 31.</b> MCGH v1: “Gene-Mutation View” .....	111
<b>Figura 32.</b> MCGH v1: “Genome View”.....	114
<b>Figura 33.</b> MCGH v1: “Transcription View” .....	116
<b>Figura 34.</b> Genotipo y Fenotipo .....	118
<b>Figura 35.</b> MCGH v1.1: “Phenotype View” .....	119
<b>Figura 36.</b> MCGH v2: “Structural View” .....	127

<b>Figura 37.</b> MCGH v2: “Transcription View” .....	130
<b>Figura 38.</b> MCGH v2: “Variation View” .....	137
<b>Figura 39.</b> “Phenotype View”: Desde versión 1.1 a versión 2 .....	143
<b>Figura 40.</b> MCGH v2: “Pathway View” .....	144
<b>Figura 41.</b> MCGH v2: “Bibliography and data bank View” .....	149
<b>Figura 42.</b> Análisis genético utilizando “variaciones” versus “variaciones + haplotipos” .....	159
<b>Figura 43.</b> Definición de haplotipos, según Sequence Ontology .....	163
<b>Figura 44.</b> Vista de Variaciones (estado actual) – Fase I.....	164
<b>Figura 45.</b> Integración de haplotipos al MCGH – Fase II .....	166
<b>Figura 46.</b> Modelo Entidad-Relación (inicial).....	172
<b>Figura 47.</b> Datos curados cargados en el repositorio de datos.....	175
<b>Figura 48.</b> Tipos de datos almacenados (total de filas) .....	175
<b>Figura 49.</b> Importación de datos utilizando HeidiSQL.....	176
<b>Figura 50.</b> Versión anterior (actual) .....	180
<b>Figura 51.</b> Nueva versión (extensión).....	181
<b>Figura 52.</b> Metodología SILE .....	189
<b>Figura 53.</b> Pantalla de bienvenida del portal de NCBI.....	190
<b>Figura 54.</b> Búsqueda de información sobre el “genoma humano” .....	190
<b>Figura 55.</b> Búsqueda del Gen “BRCA2” en el portal de NCBI .....	191
<b>Figura 56.</b> Información facilitada para el gen “BRCA2”.....	191
<b>Figura 57.</b> Búsqueda de la variación “rs671” en el portal de dbSNP.....	192
<b>Figura 58.</b> Resultado de búsqueda de la variación “rs671” en Ensembl .	192
<b>Figura 59.</b> Búsqueda de la variación “rs671” en el portal de OMIM .....	193
<b>Figura 60.</b> Búsqueda de la variación “rs671” en el portal de SNPedia ...	193
<b>Figura 61.</b> Esquema de Base de Datos (HGDB) .....	195
<b>Figura 62.</b> Propuesta carga selectiva .....	197
<b>Figura 63.</b> Módulo de carga .....	198
<b>Figura 64.</b> Trozo código Python: parser BIC.....	200
<b>Figura 65.</b> Ventana principal del prototipo Software ETL.....	200
<b>Figura 66.</b> Detalle vista de variaciones.....	201
<b>Figura 67.</b> Detalle extracción de información (Vista Estructural).....	202
<b>Figura 68.</b> Detalle Vista Fuente de datos .....	203
<b>Figura 69.</b> Estructura fichero VCF .....	204
<b>Figura 70.</b> Ejemplo fichero VCF.....	204
<b>Figura 71.</b> Aplicación VarSearch.....	206
<b>Figura 72.</b> Diagrama de Caso de Uso General: VarSearch .....	207
<b>Figura 73.</b> E-Genomic Framework y VarSearch.....	208
<b>Figura 74.</b> Arquitectura de VarSearch .....	209

<b>Figura 75.</b> Selección y carga de fichero a analizar .....	210
<b>Figura 76.</b> Tarea de análisis del fichero subido .....	210
<b>Figura 77.</b> Variaciones encontradas en la HGDB .....	211
<b>Figura 78.</b> Barra de búsqueda/filtrado.....	212
<b>Figura 79.</b> Listado de Variaciones de Usuario .....	212
<b>Figura 80.</b> Formulario de inserción de validaciones en la HGDB .....	213
<b>Figura 81.</b> Gestión de usuarios.....	214
<b>Figura 82.</b> Lista de variaciones encontradas.....	218
<b>Figura 83.</b> Lista de variaciones no encontradas .....	219
<b>Figura 84.</b> Optimización del tiempo .....	221
<b>Figura 85.</b> Diagrama de Paquete: GenesLove.Me.....	223
<b>Figura 86.</b> Arquitectura de GenesLove.Me .....	224
<b>Figura 87.</b> Página web de GenesLove.Me . .....	226

# Índice de tablas

---

<b>Tabla 1.</b> Cuatro dominios de Big Data en 2025.....	51
<b>Tabla 2.</b> Comparativa diferentes plataformas de secuenciación.....	61
<b>Tabla 3.</b> Ventajas y desventajas diferentes estrategias de secuenciación..	62
<b>Tabla 4.</b> Contenido total en la versión 70 de la base de datos COSMIC. ...	79
<b>Tabla 5.</b> Resumen Bases de Datos Genómicas .....	102
<b>Tabla 6.</b> Lista de genes y variaciones asociadas con la Sensibilidad al Alcohol .....	157
<b>Tabla 7.</b> Identificador del atributo en dbSNP .....	160
<b>Tabla 8.</b> Elementos del modelo + Fuentes de datos (origen).....	171
<b>Tabla 9.</b> Comparación entre herramientas de anotación de variantes.....	216
<b>Tabla 10.</b> Publicaciones realizadas en el marco de la Tesis Doctoral .....	235