

TABLE OF CONTENTS

	Page No.
List of Figures	i -iv
List of Tables	v-vi
List of Abbreviations	vii-ix
Supplementary Information	x
Abstract	xi-xxi

CHAPTER 1: LITERATURE REVIEW OF CNV RESEARCH

1.1	Introduction	1
1.2	Structural variations in human genome	2
1.3	Characteristics of CNV	4
1.4	Generation of CNV – Mechanisms	5
	1.4.1 Non-allelic Homologous Recombination (NAHR)	6
	1.4.2 Non Homologous End-Joining (NHEJ)	6
	1.4.3 Fork stalling and template switching (FoSTeS)	8
	1.4.4 LI Retro-transposition	9
1.5	Segmental duplications and CNV	10
1.6	Methods of detection of CNV - Algorithms and Platforms	11
	1.6.1 ArrayCGH	11
	1.6.2 BAC arrays CGH	12
	1.6.3 SNP array	12
	1.6.4 Detection of CNV using SNP array	13
	1.6.5 Challenges with SNP arrays	13
	1.6.6 CNV detection by sequencing	14
1.7	Effects of CNV – Mechanisms	15

1.8	CNV and normal phenotypic variation	16
1.9	CNVs and population genetics	20
1.10	CNVs and genome-wide association studies	22
1.11	CNV and mutation rates	25
1.12	CNV and diseases	25
	1.12.1 CNVs in infectious and autoimmune diseases	25
	1.12.2 CNVs in neurodevelopmental disorders	28
	1.12.3 CNVs in cancer	28
	1.12.4 CNVs in metabolic or cardiovascular diseases	29
1.13	CNV and its evolutionary potential	30
1.14	CNV web resources	34
	1.14.1 Common structural variation	34
	1.14.2 Clinically Significant Structural Variation	35
	1.14.3 Structural variation in cancer	35
1.15	Summary	36
1.16	Future directions	37
1.17	Objectives of the current work	38
	References	40

CHAPTER 2: LARGE COPY NUMBER VARIATION IN DIVERSE INDIAN POPULATIONS

2.1	Introduction	59
	2.1.1 CNVs and genome diversity	59
	2.1.2 Population diversity and genomic variations	60
	2.1.3 Indian populations as case study	61
	2.1.4 Large CNVs in populations	61
2.2	Objectives	62
2.3	Materials and Methodology	63
	2.3.1 Genotyping on Affymetrix 50K	63
	2.3.2 Processing of raw signal intensity data	66
	2.3.3 Detection of large CNV regions and concordance	66

	using multiple algorithms	
2.3.4	Estimation of distribution & frequency of large CNVs in different populations	67
2.3.5	Functional enrichment analysis of genes in CNV region	67
2.3.6	Validation of Selected CNV on Sequenom MassARRAY Platform	69
2.3.7	Identification of population clustering on the basis of CNVs	75
2.4	Results	76
2.4.1	Identification of large copy number variations	76
2.4.2	Concordance of CNV calls using different algorithms	80
2.4.3	Distribution and frequency among IGV populations	80
2.4.3	Identification of novel CNVs in Indian data	82
2.4.4	Functional Enrichment of CNVs	82
2.4.5	Population Stratification in Indian Data on the Basis of CNVs	83
2.4.6	Validation of CNV Calls on Sequenom MassARRAY Platforms	86
2.5	Discussion	89
2.5.1	CNV diversity in Indian populations	89
2.5.2	Functional enrichment of CNVs	90
2.5.3	Lessons from CNV validation	91
	References	93

CHAPTER 3: GENOMIC VARIATIONS IN PHENOTYPIC VARIABILITY: INSIGHTS FROM SKIN KERATINIZATION

3.1	Introduction	99
3.1.1	Skin keratinization-structure and function	99
3.1.2	Keratinization in evolutionary context	100
3.1.2	Skin keratinization and adaptation to environment	101
3.1.3	Copy number variation affecting skin keratinization	102
3.2	Objectives	103

3.3	Materials and Methodology	104
3.3.1	Identification of Gene Sets for the Analysis	104
3.3.2	Comparison of CNV prevalence in skin related pathways in Indian and worldwide populations	105
3.3.3	Association of skin related genes with environmental variables	109
3.3.4	Protein sequence divergence analysis across broad evolutionary scale	110
3.4	Results	111
3.4.1	Gene sets related to skin biological processes	111
3.4.2	Copy Number Variation in Indian and HapMap populations	112
3.4.3	Genomic variation in skin physiological processes and environmental parameters	122
3.4.5	Protein sequence divergence illustrates skin keratinization across vertebrates	128
3.5	Discussion	131
3.5.1	CNVs in skin keratinization biology	132
3.5.2	Interplay of climate and genomic variations	132
3.5.3	Protein sequence divergence in skin related processes	134
3.5.4	Evolutionary relevance of diversity in keratinization	135
3.5.5	Summary	136
	References	137

**CHAPTER 4: COPY NUMBER VARIATIONS &
SKIN PHYSIOLOGY:
LESSONS FROM INTEGRATIVE STUDY**

4.1	Introduction	145
4.1.1	Skin - Its structural details	145
4.1.2	Functional diversity of skin	146
4.1.3	Cross-talk between different skin functions	148
4.1.4	Genomic variations affecting skin phenotypes	150
4.1.5	Taking the work ahead	151

4.2	Objectives	151
4.3	Methods	152
4.3.1	Measurement of multiple phenotypic parameters related to skin physiology	152
4.3.1.1	Trans-epidermal water loss (TEWL)	155
4.3.1.2	Skin Gloss	156
4.3.1.3	Water content in skin layers	156
4.3.1.4	Elastic properties of skin	156
4.3.1.5	Melanin content and Erythema of skin	158
4.3.1.6	Sebum content	159
4.3.1.7	Skin surface pH	159
4.3.2	Analysis of multi-phenotype data for skin physiology	159
4.3.2.1	Effect of subject's age, body mass index and gender on skin phenotypes	160
4.3.2.2	Interaction between skin phenotypes	164
4.3.3	Sample Collection	165
4.3.4	DNA quantitation and QC	165
4.3.5	Genotyping using Affymetrix Genome-Wide SNP Array 6.0	165
4.3.5.1	Preparation of genomic DNA plate	167
4.3.5.2	Restriction enzyme digestion	167
4.3.5.3	Adaptor ligation	168
4.3.5.4	PCR amplification	169
4.3.5.5	PCR product purification	172
4.3.5.6	Quantitation	173
4.3.5.7	Fragmentation	173
4.3.5.8	End-labeling	175
4.3.5.9	Target hybridization	176
4.3.5.10	Washing, Staining and Scanning of arrays	177
4.3.6	Identification of Copy Number Variations	178
4.3.7	Inspection of Population stratification in the cohort	179
4.3.8	Integrative analysis of skin phenotypes and CNV data	179
4.4	Results	182

4.4.1	Analysis of multi-phenotype data for skin physiology parameters	182
4.4.1.1	Variation in elastic and aging effects in skin	182
4.4.1.2	Moisture content in different skin layers	185
4.4.1.3	Skin gloss	187
4.4.1.4	Sebum secretion from skin	190
4.4.1.5	Trans-epidermal water loss measurement in skin	191
4.4.1.6	Skin surface pH variation	192
4.4.1.7	Assessment of variation in skin pigmentation	192
4.4.2	Influence of subject's age, BMI and gender on skin phenotypes	195
4.4.2.1	Effects of subjects' age on skin related phenotypes	195
4.4.2.2	Effects of subjects' BMI on skin related phenotypes	196
4.4.3	Interaction between various skin phenotypes	204
4.4.4	Identification of CNV	208
4.4.5	Population stratification in the cohort	208
4.4.6	Integrative analysis of phenotypic data and copy number variation data	208
4.4.7	Effect of CNVs on observed phenotype values	223
4.5	Discussion	223
4.5.1	Variability in skin phenotypes	223
4.5.2	Effect of age and BMI on skin phenotypes	225
4.5.3	Skin as a system	226
4.5.4	CNVs and skin phenotypes	227
4.5.5	Summary of discussion	231
	References	232



CHAPTER 5: SUMMARY OF THESIS &		
FUTURE PROSPECTS OF CNV RESEARCH		
5.1	Summary of thesis	248
5.1.1	Overview	248
5.1.2	Summary	248
5.1.3	Conclusions	254
5.2	Challenges and future prospects of the field	255
5.2.1	Detection of CNVs in all ranges of size and in all genomic locations	256
5.2.2	Functional relevance of observed copy number variations	257
5.2.3	The connection between normal variations and CNV	257
5.2.4	Inclusion of environmental interaction to extend GWAS to GxEWAS	257
5.2.5	Selection operating on CNVs	258
5.2.6	Network approach to study genomic variations	258
5.2.7	Skin as a model system for disentangling biology	259
5.2.8	Approach to medicine for stratified populations	259
	References	261
	Appendix A: Awards and Publications	264
	Appendix B: Workshops/Conferences attended	265